Robust Spectral Compressed Sensing via Structured Matrix Completion

Yuxin Chen[†], Yuejie Chi[‡]

[†]Electrical Engineering, Stanford University
[‡]Electrical and Computer Engineering, The Ohio State University [*]

May 1, 2013

Abstract

The paper studies the problem of recovering a spectrally sparse object from a small number of time domain samples. Specifically, the object of interest with ambient dimension n is assumed to be a mixture of r complex sinusoids, while the underlying frequencies can assume any *continuous* values in the unit disk. Conventional compressed sensing paradigms suffer from the *basis mismatch* issue when imposing a discrete dictionary on the Fourier representation. To address this problem, we develop a novel nonparametric algorithm, called Enhanced Matrix Completion (EMaC), based on structured matrix completion. The algorithm starts by arranging the data into a low-rank enhanced form with multi-fold Hankel structure whose rank is upper bounded by r, and then attempts recovery via nuclear norm minimization. Under mild incoherence conditions, EMaC allows perfect recovery as soon as the number of samples exceeds the order of $r \log^2 n$, and is robust against bounded noise. Even if a constant portion of samples are corrupted with arbitrary magnitude, EMaC can still allow accurate recovery if the number of samples exceeds the order of $r \log^3 n$. Along the way, our results demonstrate that accurate completion of a low-rank multi-fold Hankel matrix is possible when the number of observed entries is proportional to the information theoretical limits (except for a logarithmic gap) under mild conditions. The performance of our algorithm and its applicability to super resolution are further demonstrated by numerical experiments.

1 Introduction

1.1 Motivation and Contributions

A large class of practical applications features high-dimensional objects that can be modeled or approximated by a superposition of spikes in the spectral (resp. time) domain, and involves estimation of the object from its time (resp. frequency) domain samples. A partial list includes acceleration of medical imaging [1], target localization in radar and sonar systems [2], inverse scattering in seismic imaging [3], fluorescence microscopy [4], channel estimation in wireless communications [5, 6], etc. The data acquisition devices, however, are often limited by hardware and physical constraints (e.g. the diffraction limit in an optical system), precluding sampling with desired resolution. It is thus of paramount interest to reduce the sensing complexity while retaining recovery resolution.

Fortunately, in many instances, it is possible to recover an object even when the number of samples is far below the ambient dimension, provided that the object has a parsimonious representation in the transform domain. In particular, recent advances in Compressed Sensing (CS) [7,8] popularize nonparametric methods based on convex surrogates. Such tractable methods do not require prior information on the model order, and are often robust against noise and outliers [9, 10].

[*]Y. Chen is with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305, USA (email: yxchen@stanford.edu). Y. Chi is with the Department of Electrical and Computer Engineering, the Ohio State University, Columbus, OH 43210, USA (email: chi@ece.osu.edu). This work will appear in part at International Conference on Machine Learning (ICML), 2013. Manuscript date: May 1, 2013.

Nevertheless, the success of CS relies on sparse representation or approximation of the object of interest in a finite discrete dictionary, while the true parameters in many applications are actually specified in a *continuous* dictionary. For concreteness, consider an object $x(\boldsymbol{t})$ that is a weighted sum of K-dimensional sinusoids at r distinct frequencies $\{\boldsymbol{f}_i \in [0,1]^K : 1 \le i \le r\}$. Conventional CS paradigms operate under the assumptions that these frequencies lie on a pre-determined grid on the unit disk. However, cautions need to be taken when imposing a discrete dictionary on continuous frequencies, since nature never poses the frequencies on the pre-determined grid, no matter how fine the grid is [11, 12]. This issue, known as *basis mismatch* between the true frequencies and the discretized grid, results in loss of sparsity due to spectral leakage along the Dirichlet kernel, and hence degeneration in the performance of CS algorithms. While one might impose finer gridding to mitigate this weakness, this approach often leads to numerical instability and high correlation between dictionary elements, which significantly weakens the advantage of these CS approaches [13].

In this paper, we consider the *spectral compressed sensing* problem, which aims to recover a spectrally sparse object from a small set of time domain samples. The underlying (possibly *multi-dimensional*) frequencies can assume any value on the unit disk, and need to be recovered with infinite precision. We develop a *nonparametric* algorithm, called Enhanced Matrix Completion (EMaC), based on structured matrix completion. Specifically, EMaC starts by converting the data samples into an enhanced matrix with K-fold Hankel structures whose rank is upper bounded by r, and then solves a nuclear norm minimization program to complete the enhanced matrix. We show that, under mild incoherence conditions, EMaC admits exact recovery from $\mathcal{O}(r \log^2 n)$ random samples[1], where r and n denote respectively the spectral sparsity and the ambient dimension. Our results further demonstrate that under the same mild incoherence conditions, EMaC is robust against bounded noise and sparse corruptions. In particular, the robust version of EMaC admits exact recovery from $\mathcal{O}(r \log^3 n)$ random samples even when a constant proportion of the samples are corrupted with arbitrary magnitudes. Furthermore, numerical experiments validate our theoretical findings, and show that EMaC is also applicable to the problem of super resolution.

Additionally, we provide theoretical guarantee for low-rank Hankel matrix completion, which is of great importance in control, natural language processing, and computer vision. To the best of our knowledge, our results provide the first theoretical bounds that are close to the information theoretic limit.

1.2 Connection to Prior Work

Spectral compressed sensing is closely related to *harmonic retrieval*, which seeks to extract the underlying frequencies of an object from a collection of its time domain samples. Conventional methods for harmonic retrieval include the Prony's method [14], ESPRIT [15], the matrix pencil method [16], the Tufts and Kumaresan approach [17], etc. These methods are typically based on the eigenvalue decomposition of covariance matrices constructed from *equi-spaced* samples, which can accommodate infinite frequency precision in the absence of noise. The K-fold Hankel structure, which plays a central role in the EMaC algorithm, roots from the traditional spectral estimation literature under the name "matrix enhancement matrix pencil" [18] for estimating multi-dimensional frequencies. One weakness of these techniques lies in that they are parametric and require prior knowledge on the model order, that is, the number of underlying frequency spikes of the signal or, at least, an estimate of it. Besides, their performance largely depends on the the knowledge of noise spectra – these methods are often sensitive to noise and unstable against outliers [19]. The finite rate of innovation approach [20, 21], while not restricting itself to equi-spaced samples, also requires prior knowledge on the model order and may not be easily extend to handle signals that are approximately sparse or vulnerable to outliers.

Nonparametric algorithms based on convex optimization differ from the above parametric techniques in that the model order does not need to be specified *a priori*. For instance, the CS paradigms [7, 8] assert that it is possible to recover a spectrally sparse signal from a small number of time domain samples. These algorithms admit faithful recovery even when the samples are contaminated by bounded noise [9, 22] or arbitrary sparse outliers [10]. By assuming the frequency spikes lie on an *a priori* grid, rather than identify the locations of frequencies, CS selects them from a pre-determined set. However, there is an inevitable *basis mismatch* between the true frequencies and the discrete grid, no matter how fine the grid is [11, 23]. A

[1] The standard notation $f(n) = \mathcal{O}(g(n))$ means that there exists a constant c such that $f(n) \le c g(n)$; $f(n) = \Theta(g(n))$ means there are constants c_1 and c_2 such that $c_1 g(n) \le f(n) \le c_2 g(n)$.

degeneration in performance of CS algorithms is reported due to loss of sparsity via spectral leakage along the Dirichlet kernel.

Recently, Candès and Fernandez-Granda [24] proposed a total-variation norm minimization algorithm to super-resolve a sparse object from frequency samples at the *low end* of the spectrum. This algorithm allows accurate super-resolution when the point sources are appropriately separated, and is stable against noise [25]. Inspired by this approach, Tang *et. al.* [13] then developed an atomic norm minimization algorithm [26] for line spectral estimation from $\mathcal{O}(r \log r \log n)$ random time domain samples, which achieves exact recovery under mild separation conditions. However, these works are limited to one-dimensional (1-D) frequency models, require that the complex signs of the frequency spikes are i.i.d. drawn from uniform distribution, and their robust against sparse outliers has not been established.

In contrast, our approach can accommodate multi-dimensional frequencies, and only assumes randomness in the observation basis. The algorithm is inspired by recent advances of Matrix Completion (MC), which aims at recovering a low-rank matrix from partial entries. It has been shown [27, 28] that exact recovery is possible via nuclear norm minimization, as soon as the number of observed entries is on the order of the information theoretic limit. This line of algorithms is also robust against noise and outliers [29, 30], and have found numerous applications in collaborative filtering [31, 32], medical imaging [33], etc. Nevertheless, the theoretical guarantees of these algorithms do not apply to the more structured observation models associated with the proposed multi-fold Hankel structure. Consequently, direct application of existing MC results yields pessimistic bounds on the required number of samples, which is far beyond the degrees of freedom underlying the sparse object.

1.3 Organization

The rest of the paper is organized as follows. The data and sample models are described in Section 2. By restricting our attention to two-dimensional (2-D) frequency models, we present the enhanced matrix and the associated matrix completion algorithms. The extension to high-dimensional frequency models is discussed in Section 2.6. The main theoretical guarantees are summarized in Sections 3, based on the incoherence conditions introduced in Section 3.1. We then discuss the extension to low-rank Hankel matrix completion in Section 4. Section 5 presents the numerical validation of our algorithms. The proofs for Theorem 1 and Theorem 3 are based on duality analysis and a golfing scheme [28], which are supplied in Section 6 and Section 7, respectively. Section 8 concludes the paper with a short summary of our findings, a discussion of potential extensions and improvements. Finally, proofs of auxiliary lemmas supporting our results are provided in the Appendix.

2 Model and Algorithm

Assume that the object of interest $x(\boldsymbol{t})$ can be modeled as a weighted sum of K-dimensional sinusoids at r distinct frequencies $\boldsymbol{f}_i \in [0,1]^K$ $(1 \le i \le r)$, i.e.

$$x(\boldsymbol{t}) = \sum_{i=1}^{r} d_i e^{j2\pi\langle \boldsymbol{t}, \boldsymbol{f}_i \rangle},$$

where d_i's denote the complex amplitude, and $\langle \cdot, \cdot \rangle$ denotes the inner product. Here, it is assumed that the frequencies $\boldsymbol{f}_i$'s are normalized with respect to the Nyquist frequency of $x(\boldsymbol{t})$. For concreteness, our discussion is mainly devoted to a 2-D frequency model. This subsumes line spectral estimation as a special case, and indicates how to address multi-dimensional models. The extension to higher dimension scenarios is discussed in Section 2.6.

2.1 2-D Frequency Model

Consider a data matrix $\boldsymbol{X} = [X_{k,l}]$, $0 \le k < n_1, 0 \le l < n_2$ of ambient dimension $n_1 \times n_2$, which can be obtained by sampling $x(\boldsymbol{t})$ on a uniform grid at the Nyquist rate. Each entry $X_{k,l}$ can be expressed as

$$X_{k,l} = \sum_{i=1}^{r} d_i y_i^k z_i^l, \tag{1}$$

where for any i $(1 \le i \le r)$, we have

$$y_i = \exp\left(j2\pi f_{1i}\right) \quad \text{and} \quad z_i = \exp\left(j2\pi f_{2i}\right),$$

for some frequency pairs $\{(f_{1i}, f_{2i}) \mid 1 \le i \le r\}$. We can then express $\boldsymbol{X}$ in a matrix form as follows

$$\boldsymbol{X} = \boldsymbol{Y}\boldsymbol{D}\boldsymbol{Z}^T, \tag{2}$$

where the above matrices are defined as

$$\boldsymbol{Y} := \begin{bmatrix} 1 & 1 & \cdots & 1 \\ y_1 & y_2 & \cdots & y_r \\ \vdots & \vdots & \vdots & \vdots \\ y_1^{n_1-1} & y_2^{n_1-1} & \cdots & y_r^{n_1-1} \end{bmatrix}, \tag{3}$$

$$\boldsymbol{Z} := \begin{bmatrix} 1 & 1 & \cdots & 1 \\ z_1 & z_2 & \cdots & z_r \\ \vdots & \vdots & \vdots & \vdots \\ z_1^{n_2-1} & z_2^{n_2-1} & \cdots & z_r^{n_2-1} \end{bmatrix}, \tag{4}$$

and

$$\boldsymbol{D} := \mathrm{diag}\left[d_1, d_2, \cdots, d_r\right]. \tag{5}$$

This is sometimes referred to as the Vandemonde decomposition of $\boldsymbol{X}$.

Suppose that there exists a location set Ω of size m such that $X_{k,l}$ is observed if and only if $(k,l) \in \Omega$. It is assume that Ω is selected uniformly at random. We are interested in recovering $\boldsymbol{X}$ from the observation of its entries on a small location set Ω.

2.2 Matrix Enhancement

One might naturally attempt recovery by applying the low-rank MC algorithms [27], arguing that when r is small, perfect recovery of $\boldsymbol{X}$ is possible from partial measurements since $\boldsymbol{X}$ is low rank if $r \ll \min\{n_1, n_2\}$. Specifically, this corresponds to the following algorithm:

$$\begin{aligned} \underset{\boldsymbol{M} \in \mathbb{C}^{n_1 \times n_2}}{\text{minimize}} \quad & \|\boldsymbol{M}\|_* \\ \text{subject to} \quad & M_{k,l} = X_{k,l}, \quad \forall\,(k,l) \in \Omega \end{aligned} \tag{6}$$

where $\|\boldsymbol{M}\|_*$ denotes the nuclear norm (or sum of all singular values) of $\boldsymbol{M} = [M_{k,l}]$. This is a convex relaxation paradigm of the rank minimization problem. However, naive MC algorithms [28] require at least the order of $r \max\left(n_1, n_2\right) \log\left(n_1 n_2\right)$ samples, which far exceeds the degrees of freedom (which is $\mathcal{O}\left(r\right)$) in our problem. What is worse, when $r > \min\left(n_1, n_2\right)$ (which is possible since r can be as large as $n_1 n_2$), $\boldsymbol{X}$ is no longer low-rank. This motivates us to construct other low-rank forms that better capture the harmonic structure.

In this paper, we adopt one effective enhanced form of $\boldsymbol{X}$ based on two-fold Hankel structure as follows. The enhanced matrix $\boldsymbol{X}_{\mathrm{e}}$ with respect to $\boldsymbol{X}$ is defined as a $k_1 \times (n_1 - k_1 + 1)$ block Hankel matrix

$$\boldsymbol{X}_{\mathrm{e}} := \begin{bmatrix} \boldsymbol{X}_0 & \boldsymbol{X}_1 & \cdots & \boldsymbol{X}_{n_1-k_1} \\ \boldsymbol{X}_1 & \boldsymbol{X}_2 & \cdots & \boldsymbol{X}_{n_1-k_1+1} \\ \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{X}_{k_1-1} & \boldsymbol{X}_{k_1} & \cdots & \boldsymbol{X}_{n_1-1} \end{bmatrix}, \tag{7}$$

where $1 \le k_1 \le n_1$, and each block is a $k_2 \times (n_2 - k_2 + 1)$ Hankel matrix defined such that for every ℓ $(0 \le \ell < n_1)$:

$$\boldsymbol{X}_\ell := \begin{bmatrix} X_{\ell,0} & X_{\ell,1} & \cdots & X_{\ell,n_2-k_2} \\ X_{\ell,1} & X_{\ell,2} & \cdots & X_{\ell,n_2-k_2+1} \\ \vdots & \vdots & \vdots & \vdots \\ X_{\ell,k_2-1} & X_{\ell,k_2} & \cdots & X_{\ell,n_2-1} \end{bmatrix}, \tag{8}$$

where $1 \le k_2 \le n_2$. This enhanced form allows us to express each block as[2]

$$\boldsymbol{X}_{\ell} = \boldsymbol{Z}_{\mathrm{L}} \boldsymbol{Y}_{\mathrm{d}}^{\ell} \boldsymbol{D} \boldsymbol{Z}_{\mathrm{R}}, \tag{9}$$

where $\boldsymbol{Z}_{\mathrm{L}}$, $\boldsymbol{Z}_{\mathrm{R}}$ and $\boldsymbol{Y}_{\mathrm{d}}$ are defined as

$$\boldsymbol{Z}_{\mathrm{L}} := \begin{bmatrix} 1 & 1 & \cdots & 1 \\ z_1 & z_2 & \cdots & z_r \\ \vdots & \vdots & \vdots & \vdots \\ z_1^{k_2-1} & z_2^{k_2-1} & \cdots & z_r^{k_2-1} \end{bmatrix},$$

$$\boldsymbol{Z}_{\mathrm{R}} := \begin{bmatrix} 1 & z_1 & \cdots & z_1^{n_2-k_2} \\ 1 & z_2 & \cdots & z_2^{n_2-k_2} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & z_r & \cdots & z_r^{n_2-k_2} \end{bmatrix},$$

and

$$\boldsymbol{Y}_{\mathrm{d}} := \operatorname{diag}\left[y_1, y_2, \cdots, y_r\right].$$

Substituting (9) into (7) yields the following:

$$\boldsymbol{X}_{\mathrm{e}} = \underbrace{\begin{bmatrix} \boldsymbol{Z}_{\mathrm{L}} \\ \boldsymbol{Z}_{\mathrm{L}} \boldsymbol{Y}_{\mathrm{d}} \\ \vdots \\ \boldsymbol{Z}_{\mathrm{L}} \boldsymbol{Y}_{\mathrm{d}}^{k_1-1} \end{bmatrix}}_{\sqrt{k_1 k_2} \boldsymbol{E}_{\mathrm{L}}} \boldsymbol{D} \underbrace{\left[\boldsymbol{Z}_{\mathrm{R}}, \quad \boldsymbol{Y}_{\mathrm{d}} \boldsymbol{Z}_{\mathrm{R}}, \quad \cdots, \quad \boldsymbol{Y}_{\mathrm{d}}^{n_1-k_1} \boldsymbol{Z}_{\mathrm{R}} \right]}_{\sqrt{(n_1-k_1+1)(n_2-k_2+1)} \boldsymbol{E}_{\mathrm{R}}}, \tag{10}$$

where $\boldsymbol{E}_{\mathrm{L}}$ and $\boldsymbol{E}_{\mathrm{R}}$ characterize the column and row space of $\boldsymbol{X}_{\mathrm{e}}$, respectively. This immediately implies that $\boldsymbol{X}_{\mathrm{e}}$ is *low-rank*, i.e.,

$$\operatorname{rank}\left(\boldsymbol{X}_{\mathrm{e}}\right) \le r. \tag{11}$$

This form aligns with the traditional matrix pencil approach proposed in [16, 18] to estimate harmonic frequencies if *all* entries of $\boldsymbol{X}$ are available. Thus, one can extract all underlying frequencies of $\boldsymbol{X}$ using methods proposed in [18], as long as $\boldsymbol{X}$ can be faithfully recovered.

2.3 The EMaC Algorithm without Noise

Define $\mathcal{P}_{\Omega}(\boldsymbol{X})$ as the orthogonal projection of $\boldsymbol{X}$ onto the subspace of matrices that vanish outside Ω. We then attempt recovery through the following Enhancement Matrix Completion (EMaC) algorithm:

$$\begin{array}{lll} \text{(EMaC)} & \underset{\boldsymbol{M} \in \mathbb{C}^{n_1 \times n_2}}{\text{minimize}} & \left\| \boldsymbol{M}_{\mathrm{e}} \right\|_* \\ & \text{subject to} & \mathcal{P}_{\Omega}\left(\boldsymbol{M}\right) = \mathcal{P}_{\Omega}\left(\boldsymbol{X}\right), \end{array} \tag{12}$$

where $\boldsymbol{M}_{\mathrm{e}}$ denotes the enhanced form of $\boldsymbol{M}$. In other words, EMaC minimizes the nuclear norm of the enhanced form over the constraint set. This convex program can be solved using off-the-shelf semidefinite program solvers in a tractable manner (see, e.g., [34]).

[2] Note that the lth row $\boldsymbol{X}_{l*}$ of $\boldsymbol{X}$ can be expressed as

$$\boldsymbol{X}_{l*} = \left[y_1^l, \cdots, y_r^l\right] \boldsymbol{D} \boldsymbol{Z} = \left[y_1^l d_1, \cdots, y_r^l d_r\right] \boldsymbol{Z},$$

and hence we only need to find the Vandemonde decomposition for $\boldsymbol{X}_0$ and then replace d_i by $y_i^l d_i$.

2.4 The Noisy-EMaC Algorithm with Bounded Noise

In practice, measurements are almost always contaminated by a certain amount of noise. To make our model and algorithm more practically applicable, we can replace our measurements by $\boldsymbol{X}^{\mathrm{o}} = [X_{kl}^{\mathrm{o}}]_{0 \le k < n_1, 0 \le l < n_2}$ through the following noisy model

$$\forall (k,l) \in \Omega : \quad X_{k,l}^{\mathrm{o}} = X_{k,l} + N_{k,l}, \tag{13}$$

where $X_{k,l}^{\mathrm{o}}$ is the observed (k,l)-th entry, and $\boldsymbol{N} = [N_{k,l}]_{0 \le k < n_1, 0 \le l < n_2}$ denotes some unknown noise. We assume that the noise magnitude is bounded by a known amount $\left\| \mathcal{P}_\Omega (\boldsymbol{N}) \right\|_{\mathrm{F}} \le \delta$, where $\|\cdot\|_{\mathrm{F}}$ denotes the Frobenius norm. In order to adapt our algorithm to such noisy measurements, one wishes that small perturbation in the measurements should result in small change in the estimate. Our algorithm is then modified as follows

$$\text{(Noisy-EMaC)} : \quad \begin{aligned} \underset{\boldsymbol{M} \in \mathbb{C}^{n_1 \times n_2}}{\text{minimize}} \quad & \|\boldsymbol{M}_{\mathrm{e}}\|_* \\ \text{subject to} \quad & \left\| \mathcal{P}_\Omega (\boldsymbol{M} - \boldsymbol{X}^{\mathrm{o}}) \right\|_{\mathrm{F}} \le \delta. \end{aligned} \tag{14}$$

That said, the algorithm searches for a candidate with minimum nuclear norm among all signals close to the measurements.

2.5 The Robust-EMaC Algorithm with Sparse Outliers

An outlier is a data sample that can deviate arbitrarily from the true data point. Practical data samples one collects may contain a certain portion of outliers due to abnormal behavior of data acquisition devices such as amplifier saturation, sensor failures, and malicious attacks. A desired recovery algorithm should be able to automatically identify the set of outliers even when they arise in up to a constant portion of all data samples.

Specifically, suppose that our measurements $\boldsymbol{X}^{\mathrm{o}}$ are given by

$$\forall (k,l) \in \Omega : \quad X_{k,l}^{\mathrm{o}} = X_{k,l} + S_{k,l}, \tag{15}$$

where $X_{k,l}^{\mathrm{o}}$ is the observed (k,l)-th entry, and $\boldsymbol{S} = [S_{k,l}]_{0 \le k < n_1, 0 \le l < n_2}$ denotes the outliers, which is assumed to be a sparse matrix supported on some location set $\Omega^{\mathrm{dirty}} \subseteq \Omega$. We assume that the sparsity pattern of $\boldsymbol{S}$ is selected uniformly at random conditioned on Ω. More specifically, we assume the following model:

1. Suppose that Ω is obtained by sampling m entries uniformly at random, and define $\rho = \frac{m}{n_1 n_2}$.

2. Conditioning on $(k,l) \in \Omega$, the events $\left\{ (k,l) \in \Omega^{\mathrm{dirty}} \right\}$ are independent with conditional probability

   $$\mathbb{P}\left\{ (k,l) \in \Omega^{\mathrm{dirty}} \mid (k,l) \in \Omega \right\} = s$$

   for some small constant corruption fraction $0 < s < 1$.

3. Define $\Omega^{\mathrm{clean}} := \Omega \backslash \Omega^{\mathrm{dirty}}$ as the location set of *uncorrupted* measurements.

EMaC is then modified as follows to accommodate sparse outliers:

$$\text{(Robust-EMaC)} \quad \begin{aligned} \underset{\boldsymbol{M}, \hat{\boldsymbol{S}} \in \mathbb{C}^{n_1 \times n_2}}{\text{minimize}} \quad & \|\boldsymbol{M}_{\mathrm{e}}\|_* + \lambda \|\hat{\boldsymbol{S}}_{\mathrm{e}}\|_1 \\ \text{subject to} \quad & \mathcal{P}_\Omega \left(\boldsymbol{M} + \hat{\boldsymbol{S}} \right) = \mathcal{P}_\Omega \left(\boldsymbol{X} + \boldsymbol{S} \right), \end{aligned} \tag{16}$$

where $\boldsymbol{M}_{\mathrm{e}}$ and $\hat{\boldsymbol{S}}_{\mathrm{e}}$ denote the enhanced form of $\boldsymbol{M}$ and $\hat{\boldsymbol{S}}$, respectively. Here, $\|\hat{\boldsymbol{S}}_{\mathrm{e}}\|_1 := \|\mathrm{vec}(\hat{\boldsymbol{S}}_{\mathrm{e}})\|_1$ denotes the ℓ_1-norm of the vectorized $\hat{\boldsymbol{S}}_{\mathrm{e}}$. The objective function of Robust-EMaC is the convex envelope of $\mathrm{rank}\left(\boldsymbol{M}_{\mathrm{e}}\right) + \lambda \|\hat{\boldsymbol{S}}_{\mathrm{e}}\|_0$, which captures the low-rankness of the enhanced form as well as the sparsity of the outliers.

2.6 Extension to Higher-Dimensional Frequency Models

The EMaC method extends to higher dimensional frequency models without difficulty. In fact, for K-dimensional frequency models, one can arrange the original data into a K-fold Hankel matrix of rank at most r. For instance, consider a 3-D model such that

$$\forall \ (l_1, l_2, l_3) \in [n_1] \times [n_2] \times [n_3] : \quad X_{l_1,l_2,l_3} = \sum_{i=1}^{r} d_i y_i^{l_1} z_i^{l_2} w_i^{l_3}.$$

An enhanced form can be defined as a 3-fold Hankel matrix such that

$$\boldsymbol{X}_{\mathrm{e}} := \begin{bmatrix} \boldsymbol{X}_{0,\mathrm{e}} & \boldsymbol{X}_{1,\mathrm{e}} & \cdots & \boldsymbol{X}_{n_3-k_3,\mathrm{e}} \\ \boldsymbol{X}_{1,\mathrm{e}} & \boldsymbol{X}_{2,\mathrm{e}} & \cdots & \boldsymbol{X}_{n_3-k_3+1,\mathrm{e}} \\ \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{X}_{k_3-1,\mathrm{e}} & \boldsymbol{X}_{k_1,\mathrm{e}} & \cdots & \boldsymbol{X}_{n_3-1,\mathrm{e}} \end{bmatrix},$$

where $\boldsymbol{X}_{i,\mathrm{e}}$ denotes the 2-D enhanced form of the matrix consisting of all entries X_{l_1,l_2,l_3} obeying $l_3 = i$. One can verify that $\boldsymbol{X}_{\mathrm{e}}$ is of rank at most r, and can thereby apply EMaC on the 3-D enhanced form. To summarize, for K-dimensional frequency models, EMaC (resp. Noisy-EMaC, Robust-EMaC) searches over all K-fold Hankel matrices that are faithful with the measurements.

2.7 Notations

Before continuing, we introduce a few notations that will be used throughout. Let the singular value decomposition (SVD) of $\boldsymbol{X}_{\mathrm{e}}$ be $\boldsymbol{X}_{\mathrm{e}} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}^*$. Denote by

$$T := \left\{ \boldsymbol{U}\boldsymbol{M}^* + \tilde{\boldsymbol{M}}\boldsymbol{V}^* : \boldsymbol{M} \in \mathbb{C}^{(n_1-k_1+1)(n_2-k_1+1)\times r}; \tilde{\boldsymbol{M}} \in \mathbb{C}^{k_1 k_2 \times r} \right\}$$

the tangent space with respect to $\boldsymbol{X}_{\mathrm{e}}$, and $T^{\perp}$ the orthogonal complement of T. Denote by $\mathcal{P}_U$ (resp. $\mathcal{P}_V$, $\mathcal{P}_T$) the orthogonal projections onto the column (resp. row, tangent) space of $\boldsymbol{X}_{\mathrm{e}}$, i.e. for any $\boldsymbol{M}$

$$\mathcal{P}_U \boldsymbol{M} = \boldsymbol{U}\boldsymbol{U}^*\boldsymbol{M}; \quad \mathcal{P}_V \boldsymbol{M} = \boldsymbol{M}\boldsymbol{V}\boldsymbol{V}^*; \quad \mathcal{P}_T = \mathcal{P}_U + \mathcal{P}_V - \mathcal{P}_U\mathcal{P}_V.$$

We let $\mathcal{P}_{T^\perp} = \mathcal{I} - \mathcal{P}_T$ be the orthogonal complement of $\mathcal{P}_T$, where $\mathcal{I}$ denotes the identity operator.

Denote by $\|\boldsymbol{M}\|$, $\|\boldsymbol{M}\|_{\mathrm{F}}$, $\|\boldsymbol{M}\|_*$ the spectral norm (operator norm), Frobenius norm, and nuclear norm of $\boldsymbol{M}$. Also, $\|\boldsymbol{M}\|_1$ and $\|\boldsymbol{M}\|_\infty$ are defined to be the ℓ_1 and ℓ_∞ norm of the vectorized $\boldsymbol{M}$. Additionally, we use $\mathrm{sgn}\,(\boldsymbol{M})$ to denote the elementwise complex sign of $\boldsymbol{M}$.

On the other hand, we denote by $\Omega_{\mathrm{e}}(k,l)$ the set of locations of the enhanced matrix $\boldsymbol{X}_{\mathrm{e}}$ containing copies of $X_{k,l}$. Due to the Hankel and multi-fold Hankel structures, one can easily verify the following: for any $\Omega_{\mathrm{e}}(k,l)$, there exists at most one index in any given row of the enhanced form, and at most one index in any given column. For each $(k,l) \in [n_1] \times [n_2]$, we use $\boldsymbol{A}_{(k,l)}$ to denote a basis matrix that extracts the average of all entries in $\Omega_{\mathrm{e}}\,(k,l)$. Specifically,

$$\left(\boldsymbol{A}_{(k,l)}\right)_{\alpha,\beta} := \begin{cases} \frac{1}{\sqrt{|\Omega_{\mathrm{e}}(k,l)|}}, & \text{if } (\alpha,\beta) \in \Omega_{\mathrm{e}}\,(k,l)\,, \\ 0, & \text{else.} \end{cases} \tag{17}$$

We will use $\omega_{k,l} := |\Omega_{\mathrm{e}}\,(k,l)|$ as a short hand notation.

3 Main Results

Encouragingly, under certain incoherence conditions, the simple EMaC enables faithful recovery of the true data matrix from a small number of noiseless time-domain samples, even when the samples are contaminated by bounded noise and a constant portion of arbitrary outliers.

3.1 Incoherence Measures

In general, matrix completion from a few entries is hopeless unless the underlying structure is uncorrelated with the observation basis. This inspires us to introduce certain incoherence measures. Let $\boldsymbol{G}_{\mathrm{L}}$ and $\boldsymbol{G}_{\mathrm{R}}$ be two $r \times r$ correlation matrices such that

$$(\boldsymbol{G}_{\mathrm{L}})_{i_1,i_2} := \begin{cases} \frac{1}{k_1 k_2} \frac{1-\left(y_{i_1}^* y_{i_2}\right)^{k_1}}{1-y_{i_1}^* y_{i_2}} \frac{1-\left(z_{i_1}^* z_{i_2}\right)^{k_2}}{1-z_{i_1}^* z_{i_2}}, & \text{if } i_1 \neq i_2, \\ 1, & \text{if } i_1 = i_2, \end{cases}$$

$$(\boldsymbol{G}_{\mathrm{R}})_{i_1,i_2} := \begin{cases} \frac{1}{(n_1-k_1+1)(n_2-k_2+1)} \frac{1-\left(y_{i_1}^* y_{i_2}\right)^{n_1-k_1+1}}{1-y_{i_1}^* y_{i_2}} \frac{1-\left(z_{i_1}^* z_{i_2}\right)^{n_2-k_2+1}}{1-z_{i_1}^* z_{i_2}}, & \text{if } i_1 \neq i_2, \\ 1, & \text{if } i_1 = i_2. \end{cases}$$

Denote the smallest singular values of $\boldsymbol{G}_{\mathrm{L}}$ and $\boldsymbol{G}_{\mathrm{R}}$ as $\sigma_{\min}(\boldsymbol{G}_{\mathrm{L}})$ and $\sigma_{\min}(\boldsymbol{G}_{\mathrm{R}})$. Note that $\boldsymbol{G}_{\mathrm{L}}$ and $\boldsymbol{G}_{\mathrm{R}}$ can be obtained by sampling the 2-D Dirichlet kernel, which is frequently considered in Fourier analysis. Our incoherence measure is then defined as follows.

Definition 1. ***[Incoherence]*** *Let $\boldsymbol{X}_{\mathrm{e}}$ denote the enhanced matrix associated with $\boldsymbol{X}$, and suppose the SVD of $\boldsymbol{X}_{\mathrm{e}}$ is given by $\boldsymbol{X}_{\mathrm{e}} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}^*$. Then $\boldsymbol{X}$ is said to have incoherence (μ_1, μ_2, μ_3) if they are respectively the smallest values obeying*

$$\sigma_{\min}(\boldsymbol{G}_L) \geq \frac{1}{\mu_1}, \quad \sigma_{\min}(\boldsymbol{G}_R) \geq \frac{1}{\mu_1}; \tag{18}$$

$$\max_{(k,l)\in[n_1]\times[n_2]} \frac{1}{\omega_{k,l}^2} \left| \sum_{(\alpha,\beta)\in\Omega_{\mathrm{e}}(k,l)} (\boldsymbol{U}\boldsymbol{V}^*)_{\alpha,\beta} \right|^2 \leq \frac{\mu_2 r}{n_1^2 n_2^2}; \tag{19}$$

$$\forall (k,l) \in [n_1]\times[n_2]: \sum_{(\alpha,\beta)\in[n_1]\times[n_2]} \omega_{\alpha,\beta} \left| \left\langle \boldsymbol{U}\boldsymbol{U}^* \boldsymbol{A}_{(k,l)} \boldsymbol{V}\boldsymbol{V}^*, \boldsymbol{A}_{(\alpha,\beta)} \right\rangle \right|^2 \leq \frac{\mu_3 r}{n_1 n_2} \omega_{k,l}. \tag{20}$$

There is a large class of spectrally sparse 2-D signals that exhibit good incoherence properties. In this paper, we will always choose $k_1 = \Theta(n_1)$, $k_2 = \Theta(n_2)$, $n_1 - k_1 = \Theta(n_1)$, and $n_2 - k_2 = \Theta(n_1)$. To give the reader a flavor for the incoherence conditions, some interpretations are in order as well as a few examples. In this subsection, we restrict our attention to square matrices, i.e. $n := n_1 = n_2$. Note that the class of incoherent signals are far beyond the ones discussed below.

3.1.1 Condition (18)

This specifies certain incoherence among the locations of frequency pairs, which does not coincide with and is not subsumed by the separation condition required in [13, 24]. The frequencies satisfying Condition (18) can be either spread out or minimally separated. Two examples are listed as follows.

- *Random frequency locations*: suppose that the r frequencies are generated uniformly at random, then the minimum pairwise separation can be crudely bounded by $\Theta\left(\frac{1}{r^2 \log n}\right)$. If $n \gg r^{2.5}\log n$, then a crude bound yields

$$\forall i_1 \neq i_2, \quad \max\left\{ \frac{1}{k_1} \frac{1-\left(y_{i_1}^* y_{i_2}\right)^{k_1}}{1-y_{i_1}^* y_{i_2}}, \frac{1}{k_2} \frac{1-\left(z_{i_1}^* z_{i_2}\right)^{k_2}}{1-z_{i_1}^* z_{i_2}} \right\} \ll \frac{1}{\sqrt{r}},$$

  indicating that the off-diagonal entries of $\boldsymbol{G}_{\mathrm{L}}$ and $\boldsymbol{G}_{\mathrm{R}}$ are much smaller than $1/r$ in magnitude. Simple manipulation then allows us to conclude that $\sigma_{\min}(\boldsymbol{G}_{\mathrm{L}})$ and $\sigma_{\min}(\boldsymbol{G}_{\mathrm{R}})$ are bounded below by positive constants.

- *Small perturbation off the grid*: suppose that all frequencies are within a distance at most $\frac{1}{nr^{1/4}}$ from some grid points $\left(\frac{l_1}{n},\frac{l_2}{n}\right)$ $(0\leq l_1,l_2<n)$. One can verify that

$$\forall i_1\neq i_2,\quad \max\left\{\frac{1}{k_1}\frac{1-\left(y_{i_1}^*y_{i_2}\right)^{k_1}}{1-y_{i_1}^*y_{i_2}},\frac{1}{k_2}\frac{1-\left(z_{i_1}^*z_{i_2}\right)^{k_2}}{1-z_{i_1}^*z_{i_2}}\right\}<\frac{1}{2\sqrt{r}},$$

  and hence the magnitude of all off-diagonal entries of $\boldsymbol{G}_{\mathrm{L}}$ and $\boldsymbol{G}_{\mathrm{R}}$ are no larger than $1/(4r)$. This immediately reveals that $\sigma_{\min}\left(\boldsymbol{G}_{\mathrm{L}}\right)$ and $\sigma_{\min}\left(\boldsymbol{G}_{\mathrm{R}}\right)$ are lower bounded by $3/4$.

3.1.2 Condition (19)

This can be satisfied when the total energy of each skew diagonal of $\boldsymbol{UV}^*$ is proportional to the dimension of this skew diagonal. This is weaker than the one introduced in [27] for matrix completion, which requires uniform energy distribution over all entries of $\boldsymbol{UV}^*$. In general, the matrix $\boldsymbol{UV}^*$ relies jointly on the frequencies and their complex amplitudes. For instance, an ideal μ_2 may arises when the complex phases of all frequencies are generated in some random fashion.

3.1.3 Condition (20)

This is an incoherence measure based on the (K-fold) Hankel structure, which depends on the energy distribution between $\mathcal{A}_{(k,l)}$ and the column and row spaces $\boldsymbol{U},\boldsymbol{V}$. The following manipulation may allow the reader to get a flavor for when this incoherence condition may hold.

Let $\boldsymbol{a},\boldsymbol{b}\in[n_1]\times[n_2]$. One can easily verify that

$$\left|\langle\boldsymbol{A}_{\boldsymbol{b}},\boldsymbol{UU}^*\boldsymbol{A}_{\boldsymbol{a}}\boldsymbol{VV}^*\rangle\right|\leq\sqrt{\omega_{\boldsymbol{b}}}\max_{(\alpha,\beta)\in\Omega_{\mathrm{e}}(\boldsymbol{b})}\left|\left(\boldsymbol{UU}^*\boldsymbol{A}_{\boldsymbol{a}}\boldsymbol{VV}^*\right)_{\alpha,\beta}\right|,$$

which, through simple manipulation, leads to

$$\sum_{\boldsymbol{a}\in[n_1]\times[n_2]}\left|\langle\boldsymbol{UU}^*\boldsymbol{A}_{\boldsymbol{b}}\boldsymbol{VV}^*,\boldsymbol{A}_{\boldsymbol{a}}\rangle\right|^2\leq\ \omega_{\boldsymbol{b}}\sum_{\boldsymbol{a}\in[n_1]\times[n_2]}\max_{(\alpha,\beta)\in\Omega_{\mathrm{e}}(\boldsymbol{b})}\left|\left(\boldsymbol{UU}^*\boldsymbol{A}_{\boldsymbol{a}}\boldsymbol{VV}^*\right)_{\alpha,\beta}\right|^2.$$

Note that $\omega_{\boldsymbol{a}}\left\|\boldsymbol{UU}^*\boldsymbol{A}_{\boldsymbol{a}}\boldsymbol{VV}^*\right\|_{\mathrm{F}}^2\leq\left\|\boldsymbol{UU}^*\right\|_{\mathrm{F}}^2=r$. If the energy of $\boldsymbol{UU}^*\boldsymbol{A}_{\boldsymbol{a}}\boldsymbol{VV}^*$ is spread out, i.e. most entries of $\omega_{\boldsymbol{a}}\boldsymbol{UU}^*\boldsymbol{A}_{\boldsymbol{a}}\boldsymbol{VV}^*$ on $\Omega_{\mathrm{e}}(\boldsymbol{b})$ are bounded by $\mathcal{O}\left(\frac{\sqrt{r}}{n^2}\right)$, then one can bound

$$\sum_{\boldsymbol{a}\in[n_1]\times[n_2]}\omega_{\boldsymbol{a}}\left|\langle\boldsymbol{UU}^*\boldsymbol{A}_{\boldsymbol{b}}\boldsymbol{VV}^*,\boldsymbol{A}_{\boldsymbol{a}}\rangle\right|^2\leq\ \omega_{\boldsymbol{b}}\sum_{\boldsymbol{a}\in[n_1]\times[n_2]}\mathcal{O}\left(\frac{r}{n^4}\right)=\omega_{\boldsymbol{b}}\mathcal{O}\left(\frac{r}{n^2}\right).$$

In this situation, we have a good incoherence measure μ_3.

3.1.4 Relations among Conditions (18), (19) and (20)

Although it is not easy to examine Conditions (19) and (20) directly through the properties of frequency pairs, we can bound μ_2 and μ_3 using the knowledge of μ_1. Their mutual relations are summarized in the following lemma, which implies that $\mu_2=\mathcal{O}\left(\mu_1^2r\right)$ and $\mu_3=\mathcal{O}\left(\mu_1^2r\right)$.

Lemma 1. *Define*

$$c_s:=\max\left(\frac{n_1n_2}{k_1k_2},\frac{n_1n_2}{\left(n_1-k_1+1\right)\left(n_2-k_2+1\right)}\right).$$

Then the incoherence measures (μ_1,μ_2,μ_3) *of* $\boldsymbol{X}_{\mathrm{e}}$ *satisfy*

$$\mu_2\leq\mu_1^2c_{\mathrm{s}}^2r,\quad\textit{and}\quad\mu_3\leq\mu_1^2c_{\mathrm{s}}^2r.$$

Proof. See Lemma 2 in Section 6. □

3.2 Theoretical Guarantees

With the above incoherence measures, the main theoretical guarantees are supplied in the following three theorems each accounting for a distinct data model: (1) noiseless measurements, (2) measurements contaminated by bounded noise, and (3) measurements corrupted by a constant portion of arbitrary outliers.

3.2.1 Exact Recovery from Noiseless Measurements

The theorem below shows that exact recovery is possible under mild incoherence conditions.

Theorem 1. *Let $\boldsymbol{X}$ be a data matrix with matrix form (2), and Ω the random location set of size m. If all measurements are noiseless, then there exists a constant $c_1 > 0$ such that under either of the following conditions:*

*i) (**strong incoherence condition**) (18), (19) and (20) hold and*

$$m > c_1 \max\left(\mu_1 c_{\mathrm{s}}, \mu_2, \mu_3 c_{\mathrm{s}}\right) r \log^2\left(n_1 n_2\right); \tag{21}$$

*ii) (**weak incoherence condition**) (18) holds and*

$$m > c_1 \mu_1^2 c_{\mathrm{s}}^2 r^2 \log^2 (n_1 n_2); \tag{22}$$

$\boldsymbol{X}$ is the unique solution of EMaC with probability exceeding $1 - (n_1 n_2)^{-2}$.

Note that the result under conditions ii) is an immediate consequence of that under conditions i) by Lemma 1. Theorem 1 states the following: (1) under *strong* incoherence condition (i.e. given that (μ_1, μ_2, μ_3) are all constants), prefect recovery is possible as soon as the number of measurements exceeds $\mathcal{O}(r \log^2(n_1 n_2))$; (2) under *weak* incoherence condition (i.e. given only that μ_1 is a constant), exact recovery is possible from $\mathcal{O}\left(r^2 \log\left(n_1 n_2\right)\right)$ samples. Since there are at least $\Theta(r)$ degrees of freedom in total, the lower bound should be no smaller than $\Theta(r)$ [3]. This establishes the orderwise optimality of EMaC under strong incoherence condition except for a logarithmic gap.

We would like to emphasize that while we assume random observation models, the conditions imposed on the data model are deterministic. This is different from [13], where randomness are assumed for both the observation model and the data model.

3.2.2 Stable Recovery in the Presence of Bounded Noise

Our method enables stable recovery even when the time domain samples are noisy copies of the true data. Here, we say the recovery is stable if the solution of Noisy-EMaC is "close" to the ground truth. To this end, we establish the following theorem, which is a counterpart of Theorem 1 in the noisy setting.

Theorem 2. *Suppose $\boldsymbol{X}^{\mathrm{o}}$ is a noisy copy of $\boldsymbol{X}$ that satisfies $\|\mathcal{P}_\Omega(\boldsymbol{X} - \boldsymbol{X}^{\mathrm{o}})\|_{\mathrm{F}} \leq \delta$. Under the conditions of Theorem 1, the solution to Noisy-EMaC in (14) satisfies*

$$\|\hat{\boldsymbol{X}}_{\mathrm{e}} - \boldsymbol{X}_{\mathrm{e}}\|_{\mathrm{F}} \leq \left\{ 2\sqrt{n_1 n_2} + 8 n_1 n_2 + \frac{8\sqrt{2} n_1^2 n_2^2}{m} \right\} \delta \tag{23}$$

with probability exceeding $1 - (n_1 n_2)^{-2}$.

Proof. See Appendix J. □

Theorem 2 basically implies that the recovered enhanced matrix (which contains $\Theta(n_1^2 n_2^2)$ entries) is close to the true enhanced matrix at high signal-to-noise ratio. In particular, the average entry inaccuracy is bounded above by $\mathcal{O}(\frac{n_1 n_2}{m}\delta)$. We note that in practice, Noisy-EMaC usually yields much better estimate, possibly by a polynomial factor. The practical applicability will be illustrated in Section 5 through numerical examples.

[3] In fact, the minimum number of measurements may better be thought of as $\Theta\left(r \log\left(n_1 n_2\right)\right)$ rather than $\Theta(r)$ due to a coupon collector's effect.

3.2.3 Robust Recovery in the Presence of Sparse Outliers

The theoretical performance of Robust-EMaC is summarized in the following theorem.

Theorem 3. *Let $\boldsymbol{X}$ be a data matrix with matrix form (2), and Ω a random location set of size m. Set $\lambda = \frac{1}{\sqrt{m \log(n_1 n_2)}}$, and assume s is some small positive constant. Suppose that the complex sign of $\boldsymbol{S}$ is randomly generated such that $\mathbb{E}sgn(\boldsymbol{S}) = 0$. Then there exist constants $c_1, c_2 > 0$ such that under either of the following conditions:*

*i) (**strong incoherence condition**) (18), (19) and (20) hold and*

$$m > c_1 \max(\mu_1 c_s, \mu_2, \mu_3 c_s)\, r \log^3(n_1 n_2), \tag{24}$$

*ii) (**weak incoherence condition**) (18) holds and*

$$m > c_1 \mu_1^2 c_s^2 r^2 \log^3(n_1 n_2), \tag{25}$$

then Robust-EMaC is exact, i.e. the minimizer $\left(\hat{\boldsymbol{M}}, \hat{\boldsymbol{S}}\right)$ satisfies $\hat{\boldsymbol{M}} = \boldsymbol{X}$, with probability exceeding $1 - (n_1 n_2)^{-2}$.

Theorem 3 specifies a candidate choice of regularization parameter λ that allows orderwise optimality, which only depends on the size of Ω but is otherwise data-independent. In practice, however, λ may better be selected via cross validation.

Furthermore, Theorem 3 demonstrates the possibility of robust recovery under a constant proportion of sparse corruptions. It states the following: (1) under *strong* incoherence condition, robust recovery is possible as soon as the number of measurements exceeds $\mathcal{O}\left(r \log^3(n_1 n_2)\right)$; (2) under *weak* incoherence condition, robust recovery is possible from $\mathcal{O}\left(r^2 \log(n_1 n_2)\right)$ samples. Compared with Theorem 1, an additional $\mathcal{O}(\log(n_1 n_2))$ factor occurs. We note, however, that these conditions can be refined via finer tuning of concentration of measure inequalities.

4 Structured Matrix Completion

One problem closely related to our method is completion of multi-fold Hankel matrices from a small number of entries. While each spectrally sparse signal can be mapped to a low-rank multi-fold Hankel matrix, it is not clear whether all multi-fold Hankel matrices of rank r can be written as the enhanced form of an object with spectral sparsity r. Therefore, one can think of recovery of multi-fold Hankel matrices as a more general problem than the spectral compressed sensing problem. Indeed, Hankel matrix completion has found numerous applications in system identification [35,36], natural language processing [37], computer vision [38], magnetic resonance imaging [39], etc.

There has been several work concerning algorithms and numerical experiments for Hankel matrix completions [35, 36, 40]. However, to the best of our knowledge, there has been little theoretical guarantee that addresses directly Hankel matrix completion. Our analysis framework can be straightforwardly adapted to the general K-fold Hankel matrix completions. Notice that μ_2 and μ_3 are defined using the SVD of $\boldsymbol{X}_e$ in (19) and (20), and we only need to modify the definition of μ_1, as stated in the following theorem.

Theorem 4. *Consider a K-fold Hankel matrix $\boldsymbol{X}_e$ of rank r. The bounds in Theorems 1, 2 and 3 continue to hold, if the incoherence μ_1 is defined as the smallest number that satisfies*

$$\max_{(k,l)\in[n_1]\times[n_2]} \left\{ \|\boldsymbol{U}\boldsymbol{U}^* \boldsymbol{A}_{(k,l)}\|_F^2, \|\boldsymbol{A}_{(k,l)} \boldsymbol{V}\boldsymbol{V}^*\|_F^2 \right\} \leq \frac{\mu_1 c_s r}{n_1 n_2}. \tag{26}$$

Proof. See Appendix K. ☐

Condition (26) requires that the left and right singular vectors are sufficiently uncorrelated with the observation basis. In fact, condition (26) is a much weaker assumption than (18).

It is worth mentioning that low-rank Hankel matrices can often be converted to low-rank Toeplitz counterparts. Both Hankel and Toeplitz matrices are important forms that capture the underlying harmonic structures. Our results and analysis framework extend to low-rank Toeplitz matrix completion problem without difficulty.

5 Numerical Experiments

In this section, we present numerical examples to evaluate the performance of the EMaC algorithm and its variants under different scenarios. We further examine the application of EMaC in image super resolution. Finally, we propose an extension of singular value thresholding (SVT) developed by Cai *et. al.* [41] that exploits the multi-fold Hankel structure to handle larger scale data sets.

5.1 Dirichlet Kernel

To better understand the incoherence condition set in the theorem, we define the two-dimensional Dirichlet kernel as

$$\mathcal{K}(f_1, f_2) = \frac{1}{k_1 k_2}\left(\frac{1 - e^{-j2\pi k_1 f_1}}{1 - e^{-j2\pi f_1}}\right)\left(\frac{1 - e^{-j2\pi k_2 f_2}}{1 - e^{-j2\pi f_2}}\right),$$

where $f_1, f_2 \in [-1/2, 1/2]$. Fig. 1 (a) shows the amplitude of $\mathcal{K}(f_1, f_2)$ when $k = k_1 = k_2 = 6$. The (i_1, i_2)-th entry of the matrix $\boldsymbol{G}_{\mathrm{L}}$ is then given as

$$(\boldsymbol{G}_{\mathrm{L}})_{i_1, i_2} = \mathcal{K}(y_{i_1} - y_{i_2}, z_{i_1} - z_{i_2}).$$

The values of the off-diagonal entries decay inverse proportionally to the separation between the frequencies according to the Dirichlet kernel. We numerically evaluated the minimum eigenvalue of $\boldsymbol{G}_{\mathrm{L}}$ in Fig. 1 (b) for different $k = 6, 36, 72$ when the spikes are randomly generated and the number of spikes is given as the sparsity level. As we increase the dimension k, the minimum eigenvalue of $\boldsymbol{G}_{\mathrm{L}}$ is closer to one, verifying our theoretical argument in Section 3.1.

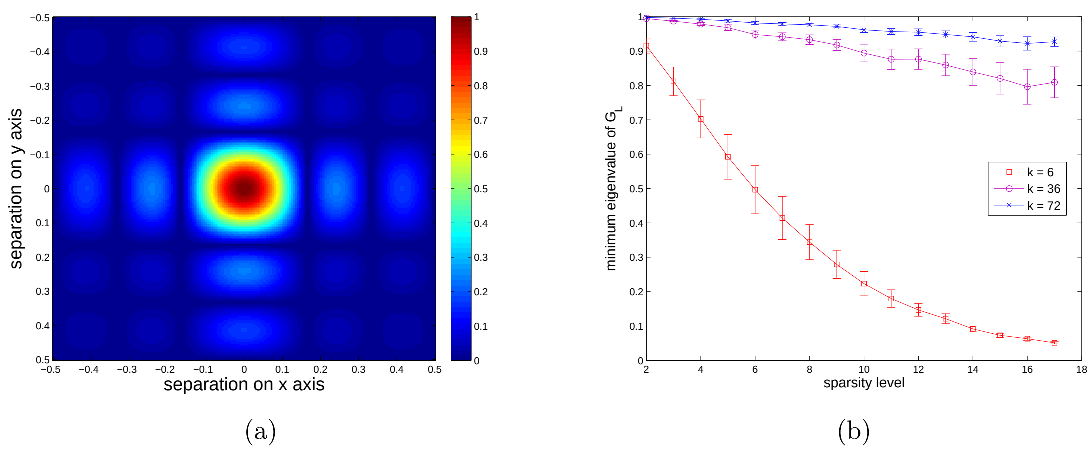
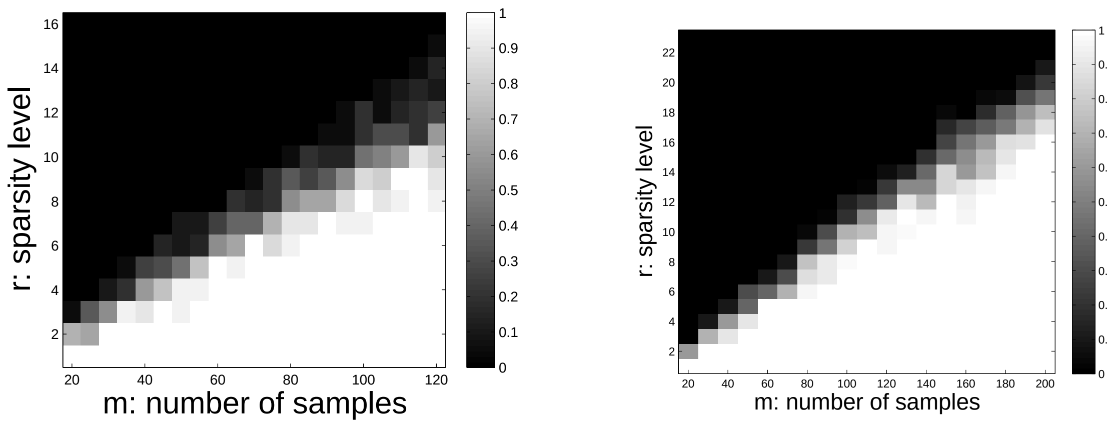


Figure 1: (a) The two-dimensional Dirichlet kernel when $k = k_1 = k_2 = 6$; (b) The empirical distribution of the minimum eigenvalue $\sigma_{\min}(\boldsymbol{G}_{\mathrm{L}})$ for different k with respect to the sparsity level.

5.2 Phase Transition in the Noiseless Setting

To evaluate the practical ability of the EMaC algorithm, we conducted a series of numerical experiments to examine the phase transition for exact recovery. A square enhanced form was adopted with $n_1 = n_2$, which corresponds to the smallest c_{s}. For each (r, m) pair, 100 Monte Carlo trials were conducted. We generated a spectrally sparse data matrix $\boldsymbol{X}$ by randomly generating r frequency spikes in $[0, 1] \times [0, 1]$, and sampled a subset Ω of size m entries uniformly at random. The EMaC algorithm was conducted using the convex programming modeling software CVX with the interior-point solver SDPT3 [42]. Each trial is declared successful if the normalized mean squared error (NMSE) satisfies $\|\hat{\boldsymbol{X}} - \boldsymbol{X}\|_{\mathrm{F}} / \|\boldsymbol{X}\|_{\mathrm{F}} \leq 10^{-3}$, where $\hat{\boldsymbol{X}}$ denotes the estimate obtained through EMaC. The empirical success rate is calculated by averaging over 100 Monte Carlo trials.

Fig. 2 illustrates the results of these Monte Carlo experiments when the dimensions of $\boldsymbol{X}$ are 11×11 and 15×15. The horizontal axis corresponds to the number m of samples revealed to the algorithm, while the vertical axis corresponds to the spectral sparsity level r. The empirical success rate is reflected by the color of each cell. It can be seen from the plot that the number of samples m grows approximately linearly with respect to the spectral sparsity r, and that the slopes of the phase transition lines for two cases are approximately the same. These observation are in line with our theoretical guarantee in Theorem 1. This phase transition diagram validates the practical applicability of our algorithm in the noiseless setting.



(a) (b)

Figure 2: Phase transition plots where frequency locations are randomly generated. The plot (a) concerns the case where $n_1 = n_2 = 11$, whereas the plot (b) corresponds to the situation where $n_1 = n_2 = 15$. The empirical success rate is calculated by averaging over 100 Monte Carlo trials.

5.3 Stable Recovery from Noisy Data

Fig. 3 further examines the stability of the proposed algorithm by performing Noisy-EMaC with respect to different parameter δ on a noise-free dataset of $r = 4$ complex sinusoids with $n_1 = n_2 = 11$. The number of random samples is $m = 50$. The reconstructed NMSE grows approximately linear with respect to δ, validating the stability of the proposed algorithm.

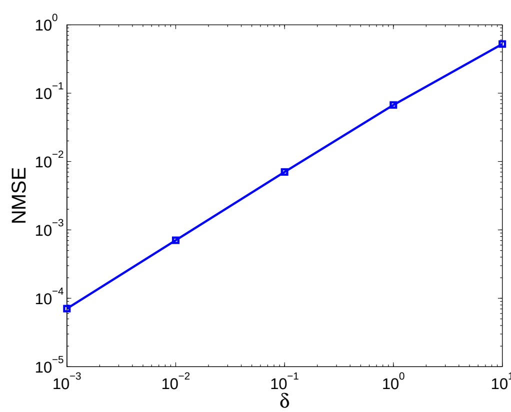
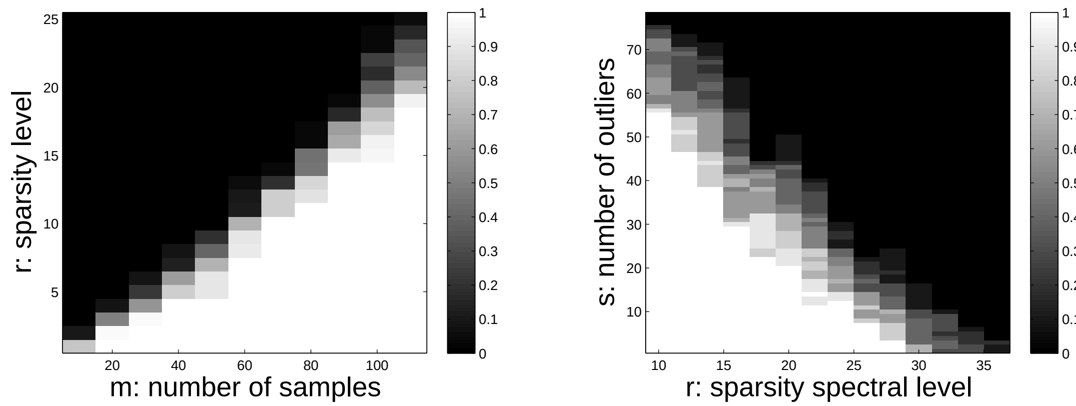


Figure 3: The reconstruction NMSE with respect to δ for a dataset with $n_1 = n_2 = 11$, $r = 4$ and $m = 50$.

5.4 Robust Line Spectrum Estimation

Consider the problem of line spectrum estimation, where the time domain measurements are contaminated by a constant portion of outliers. We conducted a series of Monte Carlo trials to illustrate the phase transition for perfect recovery of the ground truth. The true data $\boldsymbol{X}$ is assumed to be a 125-dimensional vector, where the locations of the underlying frequencies are randomly generated. The simulations were carried out again using CVX with SDPT3.

Fig. 4(a) shows the phase transition for robust line spectrum estimation when 10% of the entries are corrupted, which showcases the tradeoff between the number m of measurements and the recoverable spectral sparsity level r. One can see from the plot that m is approximately linear in r on the phase transition curve even when 10% of the measurements are corrupted, which validates our finding in Theorem 3. Fig. 4(b) illustrates the success rate of exact recovery when we obtain samples for all entry locations. This plot illustrates the tradeoff between the spectral sparsity level and the number of outliers when all entries of the corrupted $\boldsymbol{X}^o$ are observed. It can be seen that there is a large region where exact recovery can be guaranteed, demonstrating the power of our algorithms in the presence of sparse outliers.



(a) Ground truth and observation

(b) Recovered signal and corruptions

Figure 4: Robust line spectrum estimation where mode locations are randomly generated: (a) Phase transition plots when $n = 125$, and 10% of the entries are corrupted; the empirical success rate is calculated by averaging over 100 Monte Carlo trials. (b) Phase transition plots when $n = 125$, and all the entries are observed; the empirical success rate is calculated by averaging over 20 Monte Carlo trials.

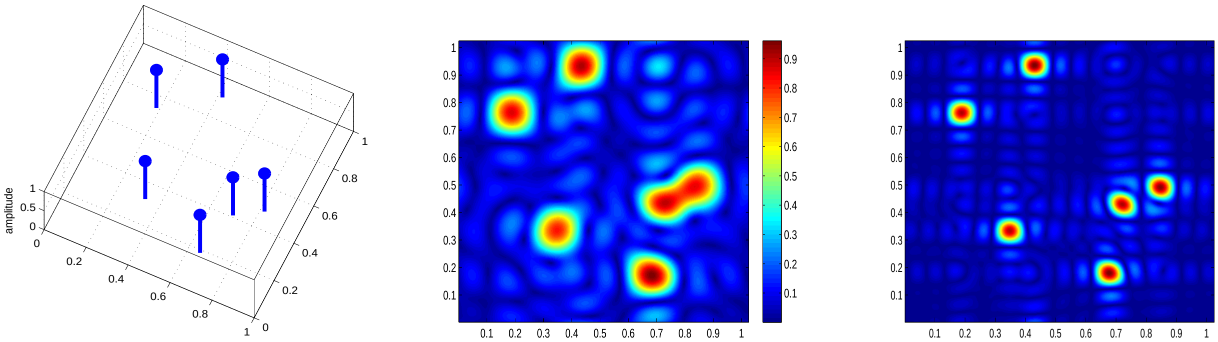
5.5 Synthetic Super Resolution

The proposed EMaC algorithm works beyond the random observation model in Theorem 1. Fig. 5 considers a synthetic super resolution example motivated by [24], where the ground truth in Fig. 5(a) contains 6 point sources with constant amplitude. The low-resolution observation in Fig. 5(b) is obtained by measuring low-frequency components $[-f_{\rm lo}, f_{\rm lo}]$ of the ground truth. Due to the large width of the associated point-spread function, both the locations and amplitudes of the point sources are distorted in the low-resolution image.

We apply EMaC to extrapolate high-frequency components up to $[-f_{\rm hi}, f_{\rm hi}]$, where $f_{\rm hi}/f_{\rm lo} = 2$. The reconstruction in Fig. 5(c) is obtained via applying directly inverse Fourier transform of the spectrum to avoid parameter estimation such as the number of modes. The resolution is greatly enhanced from Fig. 5(b), suggesting that EMaC is a promising approach for super resolution tasks.

5.6 Singular Value Thresholding for EMaC

The above Monte Carlo experiments were conducted using the advanced semidefinite programming solver SDPT3. This and many other popular solvers (e.g. SeDuMi) are based on interior point methods, which are typically inapplicable to large-scale data. In fact, SDPT3 fails to handle an $n \times n$ data matrix when n exceeds 19, which corresponds to a 100×100 enhanced matrix.



(a) Ground truth (b) Low-resolution observation (c) High-resolution reconstruction

Figure 5: A synthetic super resolution example, where the observation (b) is taken from the low-frequency components of the ground truth in (a), and the reconstruction (c) is done via inverse Fourier transform of the extrapolated high-frequency components.

One alternative for large-scale data is the first-order algorithms tailored for matrix completion problems, e.g. the singular value thresholding (SVT) algorithm [41]. We propose a modified SVT algorithm in Algorithm 1 to exploit the Hankel structure.

Algorithm 1 Singular Value Thresholding for EMaC.

Input: The observed data matrix $\boldsymbol{X}^{\mathrm{o}}$ on the location set Ω.
initialize: let $\boldsymbol{X}_{\mathrm{e}}^{\mathrm{o}}$ denote the enhanced form of $\mathcal{P}_{\Omega}\left(\boldsymbol{X}^{\mathrm{o}}\right)$; set $\boldsymbol{M}_0 = \boldsymbol{X}_{\mathrm{e}}^{0}$ and $t = 0$.
repeat
  1) $\boldsymbol{Q}_t \leftarrow \mathcal{D}_{\tau_t}\left(\boldsymbol{M}_t\right)$
  2) $\boldsymbol{M}_t \leftarrow \mathcal{H}_{\boldsymbol{X}^{\mathrm{o}}}\left(\boldsymbol{Q}_t\right)$
  3) $t \leftarrow t+1$
until convergence
output $\hat{\boldsymbol{X}}$ as the data matrix with enhanced form $\boldsymbol{M}_t$.

In particular, two operators are defined as follows:

- $\mathcal{D}_{\tau_t}(\cdot)$ in Algorithm 1 denotes the singular value shrinkage operator. Specifically, if the SVD of $\boldsymbol{X}$ is given by $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^*$ with $\boldsymbol{\Sigma} = \mathrm{diag}\left(\{\sigma_i\}\right)$, then

$$\mathcal{D}_{\tau_t}\left(\boldsymbol{X}\right) := \boldsymbol{U}\mathrm{diag}\left(\left\{(\sigma_i - \tau_t)_+\right\}\right)\boldsymbol{V}^*,$$

  where $\tau_t > 0$ is the soft-thresholding level.

- In the K-dimensional frequency model, $\mathcal{H}_{\boldsymbol{X}^{\mathrm{o}}}(\boldsymbol{Q}_t)$ denotes the projection of $\boldsymbol{Q}_t$ onto the subspace of enhanced matrices (i.e. K-fold Hankel matrices) that are consistent with the observed entries.

Consequently, at each iteration, a pair $(\boldsymbol{Q}_t, \boldsymbol{M}_t)$ is produced by first performing singular value shrinkage and then projecting the outcome onto the space of K-fold Hankel matrices that are consistent with observed entries.

Fig. 6 illustrates the performance of Algorithm 1. We generated a true 101×101 data matrix $\boldsymbol{X}$ through a superposition of 30 random complex sinusoids, and revealed 5.8% of the total entries (i.e. $m = 600$) uniformly at random. The noise was i.i.d. Gaussian giving a signal-to-noise amplitude ratio of 10. The reconstructed vectorized signal is superimposed on the ground truth in Fig. 6. The normalized reconstruction error was $\left\|\hat{\boldsymbol{X}} - \boldsymbol{X}\right\|_{\mathrm{F}} / \left\|\boldsymbol{X}\right\|_{\mathrm{F}} = 0.1098$, validating the stability of our algorithm in the presence of noise.

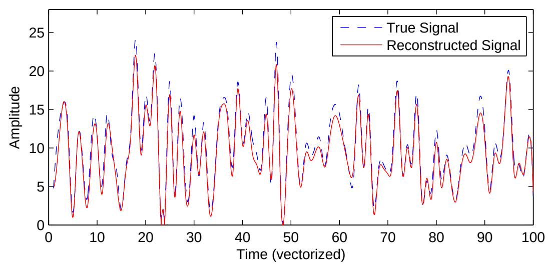


Figure 6: The performance of SVT for Noisy-EMaC for a 101×101 data matrix that contains 30 random frequency spikes. 5.8% of all entries ($m = 600$) are observed with signal-to-noise amplitude ratio 10. Here, $\tau_t = 0.1\sigma_{\max}(\boldsymbol{M}_t) / \left\lceil \frac{t}{10} \right\rceil$ empirically. For concreteness, the reconstructed data against the true data for the first 100 time instances (after vectorization) are plotted.

6 Proof of Theorem 1

EMaC has similar spirit as the well-known matrix completion algorithms [27,28] except that we impose Hankel and multi-fold Hankel structures on the matrices. While [28] has presented a general sufficient condition for exact recovery (see [28, Theorem 3]), the basis in our case does not exhibit a good coherence property as required in [28], and hence these results cannot yield useful estimates in our framework. Nevertheless, the beautiful golfing scheme introduced in [28] lays the foundation of our analysis in the sequel.

For concreteness, the analysis in this paper focuses on recovering harmonically sparse signals as stated in Theorem 1, since proving Theorem 1 is slightly more involved than proving Theorem 4. We note, however, that our analysis already entails all reasoning required for Theorem 4. In this section, we restrict our attention to the real case (i.e. $\boldsymbol{X}$ is real-valued) for simplicity[4].

Before proceeding to the proof, we would first like to stress that the incoherence measure (μ_1, μ_2, μ_3) are not independent. In addition to (μ_1, μ_2, μ_3), we define another measure μ_4 as the smallest number that satisfies

$$\forall \boldsymbol{b} \in [n_1] \times [n_2] : \quad \sum_{\boldsymbol{a} \in [n_1] \times [n_2]} \omega_{\boldsymbol{a}} \left| \langle \mathcal{P}_T \boldsymbol{A}_{\boldsymbol{b}}, \boldsymbol{A}_{\boldsymbol{a}} \rangle \right|^2 \leq \frac{\mu_4 r}{n_1 n_2} \omega_{\boldsymbol{b}}, \tag{27}$$

Some of their mutual connections are listed as follows.

Lemma 2. *Suppose that $\boldsymbol{X}_{\mathrm{e}}$ has incoherence $(\mu_1, \mu_2, \mu_3, \mu_4)$. We have the following.*

1. $\boldsymbol{G}_{\mathrm{L}} = \boldsymbol{E}_{\mathrm{L}}^* \boldsymbol{E}_{\mathrm{L}}$, *and* $\boldsymbol{G}_{\mathrm{R}} = \left(\boldsymbol{E}_{\mathrm{R}} \boldsymbol{E}_{\mathrm{R}}^*\right)^T$;

2. *For any* $\boldsymbol{a}, \boldsymbol{b} \in [n_1] \times [n_2]$, *one has*

$$\left| \langle \boldsymbol{A}_{\boldsymbol{b}}, \mathcal{P}_T \boldsymbol{A}_{\boldsymbol{a}} \rangle \right| \leq \sqrt{\frac{\omega_{\boldsymbol{b}}}{\omega_{\boldsymbol{a}}}} \frac{3\mu_1 c_{\mathrm{s}} r}{n_1 n_2}; \tag{28}$$

3. *The incoherence measure satisfies*

$$\mu_2 \leq \mu_1^2 c_{\mathrm{s}}^2 r, \quad \mu_3 \leq \mu_1^2 c_s^2 r, \tag{29}$$

and

$$\mu_4 \leq 9 \mu_1^2 c_{\mathrm{s}}^2 r; \tag{30}$$

4. *The measure μ_4 can be bounded by μ_1 and μ_3 as follows*

$$\mu_4 \leq 6\mu_1 c_{\mathrm{s}} + 3\mu_3 c_{\mathrm{s}}.$$

[4] All of the proof arguments in this paper work for Hermitian structured matrices (e.g. Hermitian Toeplitz matrices), and apply also to non-Hermitian complex case with some minor adjustment in the constants (see, e.g., [28, Section III.D]).

Proof. See Appendix A. □

Note that the above lemma indicates that our new incoherence measure μ_4 can be bounded by the sum of μ_1 and μ_3 up to some multiplicative constant. In fact, we will prove instead the following theorem based on (μ_1, μ_2, μ_4), which is slightly more general than Theorem 1.

Theorem 5. *Suppose that* $\boldsymbol{X}$ *has incoherence measure* $(\mu_1, \mu_2, \mu_3, \mu_4)$*. If*

$$m > c_0 \max\left(\mu_1 c_{\mathrm{s}}, \mu_2, \mu_4\right) r \log^2\left(n_1 n_2\right),$$

then $\boldsymbol{X}$ *is the unique solution of EMaC with probability exceeding* $1 - (n_1 n_2)^{-2}$*.*

Note that Theorem 1 can be delivered as an immediate consequence of Theorem 5 by exploiting the relations among $(\mu_1, \mu_2, \mu_3, \mu_4)$ given in Lemma 2.

6.1 Dual Certification

Denote by $\mathcal{A}_{(k,l)}\left(\boldsymbol{M}\right)$ the projection of $\boldsymbol{M}$ onto the subspace spanned by $\boldsymbol{A}_{(k,l)}$, and define the projection operator onto the space spanned by all $\boldsymbol{A}_{(k,l)}$ and its orthogonal complement as

$$\mathcal{A} := \sum_{(k,l)\in[n_1]\times[n_2]} \mathcal{A}_{(k,l)}, \quad \text{and} \quad \mathcal{A}^{\perp} = \mathcal{I} - \mathcal{A}. \tag{31}$$

Here, $\left\{\mathcal{A}^{\perp}\left(\boldsymbol{M}\right)\right\}$ spans a $\left[k_1 k_2 \left(n_1 - k_1 + 1\right)\left(n_2 - k_2 + 1\right) - n_1 n_2\right]$ dimensional subspace.

There are two common ways to describe the randomness of Ω: one corresponds to sampling *without* replacement, and another concerns sampling *with* replacement (i.e. Ω contains m indices $\{\boldsymbol{a}_i \in [n_1] \times [n_2] : 1 \leq i \leq m\}$ that are i.i.d. generated). As discussed in [28, Section II.A], while both situations result in the same order-wide bounds, the latter situation admits simpler analysis due to independence. Therefore, we will assume that Ω is a multiset (possibly with repeated elements) and a_i's are independently and uniformly distributed throughout the proofs of this paper, and define the associated operators as

$$\mathcal{A}_{\Omega} := \sum_{i=1}^{m} \mathcal{A}_{\boldsymbol{a}_i}. \tag{32}$$

We also define another projection operator $\mathcal{A}'_{\Omega}$ similar to (32), but with the sum extending only over *distinct* samples. Its complement operator is defined as $\mathcal{A}'_{\Omega^{\perp}} := \mathcal{A} - \mathcal{A}'_{\Omega}$. Note that $\mathcal{A}_{\Omega}\left(\boldsymbol{M}\right) = 0$ is equivalent to $\mathcal{A}'_{\Omega}(\boldsymbol{M}) = 0$.

With these definitions, EMaC can be rewritten as the following general matrix completion problem:

$$\begin{aligned} \underset{\boldsymbol{M}}{\text{minimize}} \quad & \left\|\boldsymbol{M}\right\|_{*} \\ \text{subject to} \quad & \mathcal{A}'_{\Omega}\left(\boldsymbol{M}\right) = \mathcal{A}'_{\Omega}\left(\boldsymbol{X}_{\mathrm{e}}\right), \\ & \mathcal{A}^{\perp}\left(\boldsymbol{M}\right) = \mathcal{A}^{\perp}\left(\boldsymbol{X}_{\mathrm{e}}\right) = 0. \end{aligned} \tag{33}$$

To prove exact recovery of convex optimization, it suffices to produce an appropriate dual certificate, as stated in the following lemma.

Lemma 3. *For a location set* Ω *that contains* m *random indices. Suppose that the sampling operator* $\mathcal{A}_{\Omega}$ *obeys*

$$\left\|\mathcal{P}_T \mathcal{A} \mathcal{P}_T - \frac{n_1 n_2}{m} \mathcal{P}_T \mathcal{A}_{\Omega} \mathcal{P}_T\right\| \leq \frac{1}{2}. \tag{34}$$

If there exists a matrix $\boldsymbol{W}$ *that obeys*

$$\mathcal{A}'_{\Omega^{\perp}}\left(\boldsymbol{U}\boldsymbol{V}^{*} + \boldsymbol{W}\right) = 0, \tag{35}$$

$$\left\|\mathcal{P}_T\left(\boldsymbol{W}\right)\right\|_{\mathrm{F}} \leq \frac{1}{2 n_1^2 n_2^2}, \tag{36}$$

and

$$\left\|\mathcal{P}_{T^{\perp}}\left(\boldsymbol{W}\right)\right\| \leq \frac{1}{2}. \tag{37}$$

Then $\boldsymbol{X}_e$ *is the unique optimizer of (33) or, equivalently,* $\boldsymbol{X}$ *is the unique minimizer of EMaC.*

Proof. See Appendix B. □

Condition (34) will be analyzed in Section 6.2, while a valid certificate $\boldsymbol{W}$ will be constructed in Section 6.3. These are the objectives of the remaining part of the section.

6.2 Deviation of $\left\|\mathcal{P}_T\mathcal{A}\mathcal{P}_T - \frac{n_1 n_2}{m}\mathcal{P}_T\mathcal{A}_\Omega\mathcal{P}_T\right\|$

Lemma 3 requires that $\mathcal{A}_\Omega$ is sufficiently incoherent with respect to T. The following lemma quantifies the projection of each $\boldsymbol{A}_{(k,l)}$ onto the tangent space T.

Lemma 4. *Suppose that* (18) *holds, then*

$$\left\|\boldsymbol{U}\boldsymbol{U}^*\boldsymbol{A}_{(k,l)}\right\|_{\mathrm{F}}^2 \le \frac{\mu_1 c_{\mathrm{s}} r}{n_1 n_2}, \quad \left\|\boldsymbol{A}_{(k,l)}\boldsymbol{V}\boldsymbol{V}^*\right\|_{\mathrm{F}}^2 \le \frac{\mu_1 c_{\mathrm{s}} r}{n_1 n_2}, \quad \left\|\mathcal{P}_T\left(\boldsymbol{A}_{(k,l)}\right)\right\|_{\mathrm{F}}^2 \le \frac{2\mu_1 c_{\mathrm{s}} r}{n_1 n_2} \tag{38}$$

for all $(k,l) \in [n_1] \times [n_2]$.

Proof. See Appendix C. □

As long as (38) holds, the deviation of $\mathcal{P}_T\mathcal{A}_\Omega\mathcal{P}_T$ can be bounded reasonably well in the following lemma. This establishes Condition (34) required by Lemma 3.

Lemma 5. *Suppose that*

$$\left\|\mathcal{P}_T\left(\boldsymbol{A}_{(k,l)}\right)\right\|_{\mathrm{F}}^2 \le \frac{2\mu_1 c_{\mathrm{s}} r}{n_1 n_2},$$

for $(k,l) \in [n_1] \times [n_2]$*. Then for any small constant* $\delta \le 2$*, one has*

$$\left\|\frac{n_1 n_2}{m}\mathcal{P}_T\mathcal{A}_\Omega\mathcal{P}_T - \mathcal{P}_T\mathcal{A}\mathcal{P}_T\right\| \le \delta \tag{39}$$

with probability exceeding $1 - 2n_1 n_2 \exp\left(-\frac{\delta^2 m}{16\mu_1 c_{\mathrm{s}} r}\right)$.

Proof. See Appendix D. □

The above two lemmas taken collectively lead to the following fact: for any given constant $\epsilon < e^{-1} < \frac{1}{2}$, $\left\|\frac{n_1 n_2}{m}\mathcal{P}_T\mathcal{A}_\Omega\mathcal{P}_T - \mathcal{P}_T\mathcal{A}\mathcal{P}_T\right\| \le \epsilon$ holds with probability exceeding $1 - (n_1 n_2)^{-4}$, provided that $m > c_1\mu_1 c_{\mathrm{s}} r \log(n_1 n_2)$ for some constant $c_1 > 0$.

6.3 Construction of Dual Certificate

Now we are in a position to construct the dual certificate, for which we will employ the golfing scheme introduced in [28]. Suppose that we generate j_0 independent random location multisets Ω_i $(1 \le i \le j_0)$, each containing $\frac{m}{j_0}$ i.i.d. samples. This way the distribution of Ω is the same as $\Omega_1 \cup \Omega_2 \cup \cdots \cup \Omega_{j_0}$. Note that Ω_i's correspond to sampling *with* replacement. Let $\rho := \frac{m}{n_1 n_2}$ and $q := \frac{\rho}{j_0}$ denote the undersampling factors of Ω and Ω_i, respectively.

Consider a small constant $\epsilon < \frac{1}{e}$, and choose $j_0 := 3\log_{\frac{1}{\epsilon}} n_1 n_2$. The construction of the dual then proceeds as follows:

Construction of a dual certificate $\boldsymbol{W}$ via the golfing scheme.
1. Set $\boldsymbol{B}_0 = 0$, and $j_0 := 3\log_{\frac{1}{\epsilon}}(n_1 n_2)$.
2. For all i $(1 \le i \le j_0)$, let $\boldsymbol{B}_i = \boldsymbol{B}_{i-1} + \left(\frac{1}{q}\mathcal{A}_{\Omega_i} + \mathcal{A}^{\perp}\right)\mathcal{P}_T\left(\boldsymbol{U}\boldsymbol{V}^* - \boldsymbol{B}_{i-1}\right)$.
3. Set $\boldsymbol{W} := -\left(\boldsymbol{U}\boldsymbol{V}^* - \boldsymbol{B}_{j_0}\right)$.

We will establish that $\boldsymbol{W}$ is a valid dual certificate if we can show that $\boldsymbol{W}$ satisfies the conditions stated in Lemma 3, which we will verify step by step.

First, by construction, we have the identities

$$\left(\mathcal{A}_\Omega' + \mathcal{A}^{\perp}\right)(\boldsymbol{B}_i) = \boldsymbol{B}_i,$$

for all $1 \le i \le j_0$. Since $\boldsymbol{U}\boldsymbol{V}^* + \boldsymbol{W} = \boldsymbol{B}_{j_0}$, this validates that $\mathcal{A}'_{\Omega^\perp}(\boldsymbol{U}\boldsymbol{V}^* + \boldsymbol{W}) = 0$, as required in (35).

Secondly, if one defines the deviation of $\mathcal{P}_T \boldsymbol{B}_i$ from $\boldsymbol{U}\boldsymbol{V}^*$ as

$$\boldsymbol{F}_i := \boldsymbol{U}\boldsymbol{V}^* - \boldsymbol{B}_i,$$

and hence $\boldsymbol{W} = \boldsymbol{F}_{j_0}$, then one can verify that

$$\begin{aligned}\mathcal{P}_T(\boldsymbol{F}_i) &= \mathcal{P}_T(\boldsymbol{U}\boldsymbol{V}^*) - \mathcal{P}_T\left(\boldsymbol{B}_{i-1} + \left(\frac{1}{q}\mathcal{A}_{\Omega_i} + \mathcal{A}^\perp\right)\mathcal{P}_T(\boldsymbol{U}\boldsymbol{V}^* - \boldsymbol{B}_{i-1})\right)\\ &= \left(\mathcal{P}_T - \mathcal{P}_T\left(\frac{1}{q}\mathcal{A}_{\Omega_i} + \mathcal{A}^\perp\right)\mathcal{P}_T\right)(\boldsymbol{F}_{i-1}).\end{aligned}$$

Lemma 5 asserts the following: if $qn_1n_2 \ge c_1\mu_1 c_{\mathrm{s}} r \log(n_1n_2)$ or, equivalently, $m \ge \tilde{c}_1\mu_1 c_{\mathrm{s}} r \log^2(n_1n_2)$, then with overwhelming probability one has

$$\left\|\mathcal{P}_T - \mathcal{P}_T\left(\frac{1}{q}\mathcal{A}_{\Omega_i} + \mathcal{A}^\perp\right)\mathcal{P}_T\right\| = \left\|\mathcal{P}_T\mathcal{A}\mathcal{P}_T - \frac{1}{q}\mathcal{P}_T\mathcal{A}_{\Omega_i}\mathcal{P}_T\right\| \le \epsilon < \frac{1}{2}.$$

This allows us to bound $\|\mathcal{P}_T(\boldsymbol{F}_i)\|_{\mathrm{F}}$ as follows

$$\|\mathcal{P}_T(\boldsymbol{F}_i)\|_{\mathrm{F}} \le \epsilon^i \|\mathcal{P}_T(\boldsymbol{F}_0)\|_{\mathrm{F}} \le \epsilon^i \|\boldsymbol{U}\boldsymbol{V}^*\|_{\mathrm{F}} = \epsilon^i\sqrt{r},$$

which immediately validates Condition (36):

$$\|\mathcal{P}_T(\boldsymbol{W})\|_{\mathrm{F}} = \|\mathcal{P}_T(\boldsymbol{F}_{j_0})\|_{\mathrm{F}} \le \epsilon^{j_0}\sqrt{r} < \frac{1}{2n_1^2n_2^2}.$$

Finally, it remains to show that $\|\mathcal{P}_{T^\perp}(\boldsymbol{W})\| \le \frac{1}{2}$. For any $\boldsymbol{F} \in T$, define the following homogeneity measure

$$\nu(\boldsymbol{F}) = \max_{(k,l)\in[n_1]\times[n_2]} \frac{1}{\omega_{k,l}} \left|\left\langle \boldsymbol{A}_{(k,l)}, \boldsymbol{F}\right\rangle\right|^2, \tag{40}$$

which largely relies on the average per-entry energy in each skew diagonal. We would like to show that $\nu\left(\left(\mathcal{I} - \mathcal{P}_T\left(\frac{1}{q}\mathcal{A}_{\Omega_i} + \mathcal{A}^\perp\right)\right)\boldsymbol{F}\right) \le \frac{1}{4}\nu(\boldsymbol{F})$ with high probability. This is supplied in the following lemma.

Lemma 6. *Consider any given $\boldsymbol{F} \in T$, and suppose that* (18) *and* (27) *hold. If the following bound holds,*

$$m > c_7 \max\{\mu_4, \mu_1 c_{\mathrm{s}}\} r \log^2(n_1n_2),$$

then one has

$$\nu\left(\left(\mathcal{P}_T - \mathcal{P}_T\left(\frac{1}{q}\mathcal{A}_{\Omega_i} + \mathcal{A}^\perp\right)\mathcal{P}_T\right)\boldsymbol{F}\right) \le \frac{1}{4}\nu(\boldsymbol{F}) \tag{41}$$

for all $1 \le i \le j_0$ with probability exceeding $1 - (n_1n_2)^{-3}$.

Proof. See Appendix E. □

This lemma basically indicates that a homogeneous $\boldsymbol{F}$ with respect to the observation basis typically results in a homogeneous $\left(\mathcal{P}_T - \mathcal{P}_T\left(\frac{1}{q}\mathcal{A}_{\Omega_i} + \mathcal{A}^\perp\right)\mathcal{P}_T\right)(\boldsymbol{F})$, and hence we can hope that the homogeneity condition (19) of $\boldsymbol{F}_0$ can carry over to every $\mathcal{P}_T(\boldsymbol{F}_i)$ $(1 \le i \le j_0)$.

Observe that Condition (19) is equivalent to saying

$$\nu(\boldsymbol{F}_0) = \max_{(k,l)\in[n_1]\times[n_2]} \frac{1}{\omega_{k,l}} \left|\left\langle \boldsymbol{A}_{(k,l)}, \boldsymbol{U}\boldsymbol{V}^*\right\rangle\right|^2 = \max_{(k,l)\in[n_1]\times[n_2]} \frac{1}{\omega_{k,l}^2} \left|\sum_{(\alpha,\beta)\in\Omega_{\mathrm{e}}(k,l)} (\boldsymbol{U}\boldsymbol{V}^*)_{\alpha,\beta}\right|^2 \le \frac{\mu_2 r}{(n_1n_2)^2}.$$

One can then verify that for every i $(0 \le i \le j_0)$,

$$\nu(\mathcal{P}_T(\boldsymbol{F}_i)) \le \frac{1}{4}\nu(\mathcal{P}_T(\boldsymbol{F}_{i-1})) \le \left(\frac{1}{4}\right)^i \nu(\boldsymbol{F}_0) \le \left(\frac{1}{4}\right)^i \frac{\mu_2 r}{(n_1n_2)^2}$$

holds with high probability if $m > c_7 \max\{\mu_4, \mu_1 c_{\mathrm{s}}\} r \log^2(n_1n_2)$ for some constant $c_7 > 0$.

The following lemma then relates the homogeneity measure with $\left\|\mathcal{P}_{T^\perp}\left(\frac{1}{q}\mathcal{A}_{\Omega_i} + \mathcal{A}^\perp\right)(\boldsymbol{F}_i)\right\|$.

Lemma 7. *For any given $\boldsymbol{F} \in T$ such that $\nu(\boldsymbol{F})$. Then there exist positive constants c_8 and c_9 such that for any $t \leq \sqrt{\nu(\boldsymbol{F})} n_1 n_2$,*

$$\left\| \mathcal{P}_{T^\perp} \left(\frac{1}{q} \mathcal{A}_{\Omega_i} + \mathcal{A}^\perp \right) (\boldsymbol{F}) \right\| > t$$

holds with probability at most $c_8 \exp\left(-\frac{c_9 q t^2}{\nu(\boldsymbol{F}) n_1 n_2} \right)$.

Proof. See Appendix F. □

Since $\nu(\boldsymbol{F}_i) \leq \left(\frac{1}{4}\right)^i \frac{\mu_2 r}{n_1^2 n_2^2}$ for all $1 \leq i \leq j_0$ with high probability, then one can bound

$$\frac{\sqrt{\nu(\boldsymbol{F}_i)} n_1 n_2}{\sqrt{16 \mu_2 r}} \leq \left(\frac{1}{2}\right)^{i+2}.$$

Lemma 7 immediately yields that for all i $(0 \leq i \leq j_0)$

$$\begin{aligned} \mathbb{P}\left\{ \forall i : \left\| \mathcal{P}_{T^\perp} \left(\frac{1}{q} \mathcal{A}_{\Omega} + \mathcal{A}^\perp \right) (\boldsymbol{F}_i) \right\| \leq \left(\frac{1}{2}\right)^{i+2} \right\} &\geq \mathbb{P}\left\{ \forall i : \left\| \mathcal{P}_{T^\perp} \left(\frac{1}{q} \mathcal{A}_{\Omega} + \mathcal{A}^\perp \right) (\boldsymbol{F}_i) \right\| \leq \frac{\sqrt{\nu(\boldsymbol{F}_i)} n_1 n_2}{\sqrt{16 \mu_2 r}} \right\} \\ &\geq 1 - c_8 n_1 n_2 \exp\left(-\frac{c_9 q n_1 n_2}{16 \mu_2 r} \right) \\ &\geq 1 - c_8 \left(n_1 n_2\right)^{-4}, \end{aligned}$$

holds if $q n_1 n_2 > c_{12} \max\left(\mu_1 c_{\mathrm{s}}, \mu_4, \mu_2\right) r \log\left(n_1 n_2\right)$ for some constant $c_{12} > 0$. This is also equivalent to

$$m > c_{13} \max\left(\mu_1 c_{\mathrm{s}}, \mu_4, \mu_2\right) r \log^2\left(n_1 n_2\right)$$

for some constant $c_{13} > 0$. Under this condition, we can conclude

$$\begin{aligned} \left\| \mathcal{P}_{T^\perp}\left(\boldsymbol{W}\right) \right\| &\leq \sum_{i=0}^{j_0} \left\| \mathcal{P}_{T^\perp} \left(\frac{1}{q} \mathcal{A}_{\Omega} + \mathcal{A}^\perp \right) (\boldsymbol{F}_i) \right\| \\ &\leq \sum_{i=0}^{j_0} \left(\frac{1}{2}\right)^{i+2} < \frac{1}{2}. \end{aligned}$$

So far, we have successfully established that with high probability, $\boldsymbol{W}$ is a valid dual certificate, and hence EMaC admits perfect reconstruction of $\boldsymbol{X}$.

7 Proof of Theorem 3

The algorithm Robust-EMaC has similar spirit as the well-known robust principal component analysis [10, 29] that seeks a decomposition of low-rank plus sparse matrices, except that we impose multi-fold Hankel structures on both the low-rank and sparse matrices. Similar to the proof for Theorem 1, the proof is based on duality analysis, and we rely on the golfing scheme introduced in [28] to construct a valid dual certificate.

In this section, we prove the results for a slightly different sampling model as follows.

- The location multiset Ω^{clean} of observed uncorrupted entries is generated by sampling $(1-s)\,\rho n_1 n_2$ i.i.d. entries uniformly at random.
- The location multiset Ω of observed entries is generated by sampling $\rho n_1 n_2$ i.i.d. entries uniformly at random, with the first $(1-s)\,\rho n_1 n_2$ entries coming from Ω^{clean}.
- The location set Ω^{dirty} of observed corrupted entries is given by $\Omega' \backslash \Omega^{\mathrm{clean}'}$, where Ω' and $\Omega^{\mathrm{clean}'}$ denote the sets of distinct entry locations in Ω and Ω^{clean}, respectively.

As mentioned in the proof of Theorem 1, this slightly different sampling model, while resulting in the same order-wise bounds, significantly simplifies the analysis due to the independent assumptions.

Similar to the proof for the noiseless setting, we will prove our results for the real case under the incoherence measures (μ_1, μ_2, μ_4) as introduced in Section 6. Moreover, we will prove the theorem under a stronger randomness condition, that is, the signs of all non-zero entries of $\boldsymbol{S}$ are *independent zero-mean* random variables. Specifically, we will prove the following theorem.

Theorem 6. *Suppose that $\boldsymbol{X}$ has incoherence measure $(\mu_1, \mu_2, \mu_3, \mu_4)$, and let $\lambda = \frac{1}{\sqrt{m \log(n_1 n_2)}}$. Assume that s is some small positive constant, and that the signs of nonzero entries of $\boldsymbol{S}$ are independently generated with mean 0. If*

$$m > c_0 \max\left(\mu_1 c_{\mathrm{s}}, \mu_2, \mu_4\right) r \log^3\left(n_1 n_2\right),$$

then Robust-EMaC succeeds in recovering $\boldsymbol{X}$ with probability exceeding $1 - (n_1 n_2)^{-2}$.

A simple derandomization argument introduced in [29, Section 2.2] immediately suggests that the performance of Robust-EMaC under the fixed-sign pattern is no worse than that under the random-sign pattern with sparsity parameter $2s$. Therefore, we will focus on the random sign pattern, which are much easier to analyze.

The same argument as in Section 6 indicates that Theorem 3 can be delivered as an immediate consequence of Theorem 6.

7.1 Dual Certification

We adopt similar notations as in Section 6.1. That said, if we generate $\rho n_1 n_2$ i.i.d. entry locations $\boldsymbol{a}_i$'s uniformly at random, and let the multisets Ω and Ω^{clean} contain respectively $\{\boldsymbol{a}_i | 1 \leq i \leq \rho n_1 n_2\}$ and $\{\boldsymbol{a}_i | 1 \leq i \leq \rho(1-s) n_1 n_2\}$), then

$$\mathcal{A}_{\Omega} := \sum_{i=1}^{\rho n_1 n_2} \mathcal{A}_{\boldsymbol{a}_i}, \quad \text{and} \quad \mathcal{A}_{\Omega^{\mathrm{clean}}} := \sum_{i=1}^{\rho(1-s) n_1 n_2} \mathcal{A}_{\boldsymbol{a}_i},$$

corresponding to sampling with replacement. Besides, $\mathcal{A}'_{\Omega}$ (resp. $\mathcal{A}'_{\Omega^{\mathrm{clean}}}$) is defined similar to $\mathcal{A}_{\Omega}$ (resp. $\mathcal{A}_{\Omega^{\mathrm{clean}}}$), but with the sum extending only over *distinct* samples.

We will establish that exact recovery can be guaranteed, if we can produce a valid dual certificate as follows.

Lemma 8. *Suppose that s is some small positive constant. Suppose that the associated sampling operator $\mathcal{A}_{\Omega^{\mathrm{clean}}}$ obeys*

$$\left\| \mathcal{P}_T \mathcal{A} \mathcal{P}_T - \frac{1}{\rho(1-s)} \mathcal{P}_T \mathcal{A}_{\Omega^{\mathrm{clean}}} \mathcal{P}_T \right\| \leq \frac{1}{2}, \tag{42}$$

and

$$\left\| \mathcal{A}_{\Omega^{\mathrm{clean}}}(\boldsymbol{M}) \right\|_{\mathrm{F}} \leq 10 \log(n_1 n_2) \left\| \mathcal{A}'_{\Omega^{\mathrm{clean}}}(\boldsymbol{M}) \right\|_{\mathrm{F}}, \tag{43}$$

for any matrix $\boldsymbol{M}$. If there exist a regularization parameter λ $(0 < \lambda < 1)$ and a matrix $\boldsymbol{W}$ obeying

$$\begin{cases} \left\| \mathcal{P}_T\left(\boldsymbol{W} + \lambda \mathrm{sgn}(\boldsymbol{S}_{\mathrm{e}}) - \boldsymbol{U}\boldsymbol{V}^*\right) \right\|_{\mathrm{F}} \leq \frac{\lambda}{n_1^2 n_2^2}, \\ \left\| \mathcal{P}_{T^\perp}\left(\boldsymbol{W} + \lambda \mathrm{sgn}(\boldsymbol{S}_{\mathrm{e}})\right) \right\| \leq \frac{1}{4}, \\ \mathcal{A}'_{(\Omega^{\mathrm{clean}})^\perp}(\boldsymbol{W}) = 0, \\ \left\| \mathcal{A}'_{\Omega^{\mathrm{clean}}}(\boldsymbol{W}) \right\|_\infty \leq \frac{\lambda}{4}, \end{cases} \tag{44}$$

then Robust-EMaC is exact, i.e. the minimizer $\left(\hat{\boldsymbol{M}}, \hat{\boldsymbol{S}}\right)$ satisfies $\hat{\boldsymbol{M}} = \boldsymbol{X}$.

Proof. See Appendix G. ∎

We note that a good bound on $\left\|\mathcal{P}_T\mathcal{A}\mathcal{P}_T-\frac{1}{\rho(1-s)}\mathcal{P}_T\mathcal{A}_{\Omega^{\text{clean}}}\mathcal{P}_T\right\|$ can be found using Lemma 5. Specifically, there exists some constant $c_1>0$ such that if $\rho(1-s)n_1n_2>c_1\mu_1c_{\text{s}}r\log(n_1n_2)$, then one has

$$\left\|\mathcal{P}_T\mathcal{A}\mathcal{P}_T-\frac{1}{\rho(1-s)}\mathcal{P}_T\mathcal{A}_{\Omega^{\text{clean}}}\mathcal{P}_T\right\|\le\frac{1}{2}$$

with probability exceeding $1-(n_1n_2)^{-4}$. Besides, a simple Chernoff bound [43] indicates that with probability exceeding $1-(n_1n_2)^{-3}$, none of the entries is sampled more than $10\log(n_1n_2)$ times. Equivalently,

$$\mathbb{P}\left(\forall\boldsymbol{M}:\|\mathcal{A}_{\Omega^{\text{clean}}}(\boldsymbol{M})\|_{\text{F}}\le10\log(n_1n_2)\left\|\mathcal{A}'_{\Omega^{\text{clean}}}(\boldsymbol{M})\right\|_{\text{F}}\right)\ge1-(n_1n_2)^{-3}.$$

Our objective in the remaining part of the section is to produce a matrix $\boldsymbol{W}$ satisfying Condition (44).

7.2 Construction of Dual Certificate

Suppose that we generate j_0 independent random location multisets Ω_j^{clean}, where Ω_j^{clean} contains qn_1n_2 i.i.d. samples uniformly at random. Here, we set $q:=\frac{(1-s)\rho}{j_0}$. This way the distribution of the multiset Ω is the same as $\Omega_1^{\text{clean}}\cup\Omega_2^{\text{clean}}\cup\cdots\cup\Omega_{j_0}^{\text{clean}}$.

We now propose constructing a dual certificate $\boldsymbol{W}$ as follows:

Construction of a dual certificate $\boldsymbol{W}$ via the golfing scheme.

1. Set $\boldsymbol{F}_0=\mathcal{P}_T(\boldsymbol{U}\boldsymbol{V}^*-\lambda\text{sgn}(\boldsymbol{S}_{\text{e}}))$, and $j_0:=5\log_{\frac{1}{\epsilon}}n_1n_2$.
2. For every i $(1\le i\le j_0)$, let $\boldsymbol{F}_i:=\left(\mathcal{P}_T\mathcal{A}\mathcal{P}_T-\frac{1}{q}\mathcal{P}_T\mathcal{A}_{\Omega_i^{\text{clean}}}\mathcal{P}_T\right)\boldsymbol{F}_{i-1}$.
3. Set $\boldsymbol{W}:=\sum_{j=1}^{j_0}\left(\frac{1}{q}\mathcal{A}_{\Omega_j^{\text{clean}}}+\mathcal{A}^{\perp}\right)\boldsymbol{F}_{j-1}$.

Take $\lambda=\frac{1}{\sqrt{m\log(n_1n_2)}}$. We will justify that $\boldsymbol{W}$ is a valid dual certificate, by examining the conditions in (44) step by step.

(1) The first condition requires $\|\mathcal{P}_T(\boldsymbol{W}+\lambda\text{sgn}(\boldsymbol{S}_{\text{e}})-\boldsymbol{U}\boldsymbol{V}^*)\|_{\text{F}}$ to be reasonably small. Lemma 5 asserts that there exist some constants $c_1,\tilde{c}_1>0$ such that if $m=\rho n_1n_2>c_1\mu_1c_{\text{s}}r\log^2(n_1n_2)$ or, equivalently, $q_in_1n_2>\tilde{c}_1\mu_1c_{\text{s}}r\log^2(n_1n_2)$, then

$$\begin{aligned}\|\mathcal{P}_T\boldsymbol{F}_{j_0}\|_{\text{F}}&\le\frac{1}{\epsilon}\|\mathcal{P}_T\boldsymbol{F}_{j_0-1}\|_{\text{F}}\le\cdots\le\left(\frac{1}{\epsilon}\right)^{j_0}\|\mathcal{P}_T\boldsymbol{F}_0\|_{\text{F}}\\&\le\frac{1}{n_1^5n_2^5}\left(\|\boldsymbol{U}\boldsymbol{V}^*\|_{\text{F}}+\lambda\|\text{sgn}(\boldsymbol{S}_{\text{e}})\|_{\text{F}}\right)&&(45)\\&\le\frac{1}{n_1^5n_2^5}\left(\sqrt{r}+\lambda n_1n_2\right)<\frac{1}{n_1^5n_2^5}(n_1n_2+\lambda n_1n_2)\ll\frac{\lambda}{n_1^2n_2^2}&&(46)\end{aligned}$$

with probability exceeding $1-(n_1n_2)^{-3}$.

It remains to relate $\mathcal{P}_T(\boldsymbol{W}+\lambda\text{sgn}(\boldsymbol{S}_{\text{e}})-\boldsymbol{U}\boldsymbol{V}^*)$ with $\mathcal{P}_T\boldsymbol{F}_{j_0}$. By construction,

$$\begin{aligned}-\mathcal{P}_T(\boldsymbol{W}+\lambda\text{sgn}(\boldsymbol{S}_{\text{e}})-\boldsymbol{U}\boldsymbol{V}^*)&=\mathcal{P}_T\boldsymbol{F}_0-\sum_{i=1}^{j_0}\mathcal{P}_T\left(\frac{1}{q}\mathcal{A}_{\Omega_i^{\text{clean}}}+\mathcal{A}^{\perp}\right)\mathcal{P}_T\boldsymbol{F}_{i-1}\\&=\mathcal{P}_T\boldsymbol{F}_0-\mathcal{P}_T\left(\frac{1}{q}\mathcal{A}_{\Omega_1^{\text{clean}}}+\mathcal{A}^{\perp}\right)\mathcal{P}_T\boldsymbol{F}_0-\sum_{i=2}^{j_0}\mathcal{P}_T\left(\frac{1}{q}\mathcal{A}_{\Omega_i^{\text{clean}}}+\mathcal{A}^{\perp}\right)\boldsymbol{F}_{i-1}\\&=\mathcal{P}_T\boldsymbol{F}_1-\sum_{i=2}^{j_0}\mathcal{P}_T\left(\frac{1}{q}\mathcal{A}_{\Omega_i^{\text{clean}}}+\mathcal{A}^{\perp}\right)\boldsymbol{F}_{i-1}\\&=\cdots=\mathcal{P}_T\boldsymbol{F}_{j_0}.\end{aligned}$$

Plugging this into (46) establishes that

$$\left\|\mathcal{P}_T\left(\boldsymbol{W}+\lambda\mathrm{sgn}\left(\boldsymbol{S}_{\mathrm{e}}\right)-\boldsymbol{U}\boldsymbol{V}^*\right)\right\|_{\mathrm{F}}=\left\|\mathcal{P}_T\boldsymbol{F}_{j_0}\right\|_{\mathrm{F}}\leq\frac{\lambda}{n_1^2n_2^2}. \tag{47}$$

(2) The second component of the proof is to develop a bound on $\left\|\mathcal{P}_{T^\perp}\left(\boldsymbol{W}+\lambda\mathrm{sgn}\left(\boldsymbol{S}_{\mathrm{e}}\right)\right)\right\|$, for which we proceed by bounding $\left\|\mathcal{P}_{T^\perp}\left(\boldsymbol{W}\right)\right\|$ and $\left\|\mathcal{P}_{T^\perp}\left(\lambda\mathrm{sgn}\left(\boldsymbol{S}_{\mathrm{e}}\right)\right)\right\|$ separately. Recall that an inhomogenuity measure $\nu\left(\boldsymbol{F}\right)$ applied to a matrix $\boldsymbol{F}\in T$ is defined in (40). The following lemma develops upper bounds on $\nu\left(\boldsymbol{F}_i\right)$ for every i $(0\leq i<j_0)$.

Lemma 9. *If* $\lambda=\frac{1}{\sqrt{m\log(n_1n_2)}}$, *then there exists constants* $c_7,c_9>0$ *such that if* $m>c_7\max\left(\mu_1c_{\mathrm{s}},\mu_4\right)r\log^2\left(n_1n_2\right)$, *for any* $0\leq i\leq j_0$,

$$\nu\left(\boldsymbol{F}_i\right)\leq\frac{c_9\max\left(\mu_1c_{\mathrm{s}},\mu_2,\mu_4\right)r}{4^in_1^2n_2^2}$$

with probability exceeding $1-n_1^{-3}n_2^{-3}$.

Proof. See Appendix H. □

Now we are ready to develop an upper bound on $\left\|\mathcal{P}_{T^\perp}\left(\boldsymbol{W}\right)\right\|$. Observe from the construction procedure of $\boldsymbol{W}$ that

$$\begin{aligned}\left\|\mathcal{P}_{T^\perp}\left(\boldsymbol{W}\right)\right\|&=\left\|\sum_{i=1}^{j_0}\mathcal{P}_{T^\perp}\left(\frac{1}{q}\mathcal{A}_{\Omega_i^{\mathrm{clean}}}+\mathcal{A}^\perp\right)\boldsymbol{F}_{i-1}\right\|\\&\leq\sum_{i=1}^{j_0}\left\|\mathcal{P}_{T^\perp}\left(\frac{1}{q}\mathcal{A}_{\Omega_i^{\mathrm{clean}}}+\mathcal{A}^\perp\right)\boldsymbol{F}_{i-1}\right\|.\end{aligned} \tag{48}$$

Combining Lemma 7 and Lemma 9 with some simple manipulations yields the following: if $m>c_{13}\max\left(\mu_1c_{\mathrm{s}},\mu_2,\mu_4\right)r\log^2\left(n_1\right.$ then one has

$$\mathbb{P}\left\{\forall i:\left\|\mathcal{P}_{T^\perp}\left(\frac{1}{q}\mathcal{A}_{\Omega_i^{\mathrm{clean}}}+\mathcal{A}^\perp\right)\boldsymbol{F}_{i-1}\right\|\leq\left(\frac{1}{2}\right)^{i+4}\right\}\geq1-\frac{c_8}{n_1^4n_2^4}.$$

Consequently, (48) can be further bounded by

$$\left\|\mathcal{P}_{T^\perp}\left(\boldsymbol{W}\right)\right\|\leq\sum_{i=0}^{j_0}\left(\frac{1}{2}\right)^{i+4}<\frac{1}{8} \tag{49}$$

with probability exceeding $1-\frac{c_8}{n_1^4n_2^4}$.

We still need to bound $\left\|\mathcal{P}_{T^\perp}\left(\lambda\mathrm{sgn}\left(\boldsymbol{S}_{\mathrm{e}}\right)\right)\right\|$, which is supplied in the following lemma.

Lemma 10. *Consider* $\lambda=\frac{1}{\sqrt{m\log(n_1n_2)}}$, *and suppose that* s *is a small constant. If* $m>c_7\max\left(\mu_1c_{\mathrm{s}},\mu_4\right)r\log^2\left(n_1n_2\right)$, *we can bound*

$$\left\|\mathrm{sgn}\left(\boldsymbol{S}_{\mathrm{e}}\right)\right\|\leq\sqrt{c_{22}\rho sn_1n_2\log^{\frac{1}{2}}\left(n_1n_2\right)}\quad\text{and}\quad\left\|\mathcal{P}_{T^\perp}\left(\lambda\mathrm{sgn}\left(\boldsymbol{S}_{\mathrm{e}}\right)\right)\right\|\leq\frac{1}{8} \tag{50}$$

with probability at least $1-\left(n_1n_2\right)^{-5}$.

Proof. See Appendix I. □

Putting (49) and (50) together yields

$$\left\|\mathcal{P}_{T^\perp}\left(\boldsymbol{W}+\lambda\mathrm{sgn}\left(\boldsymbol{S}_{\mathrm{e}}\right)\right)\right\|\leq\left\|\mathcal{P}_{T^\perp}\left(\boldsymbol{W}\right)\right\|+\left\|\mathcal{P}_{T^\perp}\left(\lambda\mathrm{sgn}\left(\boldsymbol{S}_{\mathrm{e}}\right)\right)\right\|\leq\frac{1}{4}$$

with high probability.

(3) By construction, $\mathcal{A}'_{\left(\Omega^{\mathrm{clean}}\right)^\perp}\left(\boldsymbol{W}\right)=0$.

(4) The last step is to bound $\left\|\mathcal{A}'_{\Omega^{\mathrm{clean}}}\left(\boldsymbol{W}\right)\right\|_{\infty}$, which is apparently bounded above by $\left\|\mathcal{A}_{\Omega^{\mathrm{clean}}}\left(\boldsymbol{W}\right)\right\|_{\infty}$. By construction, one can express

$$\begin{aligned}
\left\|\mathcal{A}_{\Omega^{\mathrm{clean}}}\left(\boldsymbol{W}\right)\right\|_{\infty} &= \left\|\sum_{i=1}^{j_0}\mathcal{A}_{\Omega^{\mathrm{clean}}}\left(\frac{1}{q}\mathcal{A}_{\Omega_i^{\mathrm{clean}}}+\mathcal{A}^{\perp}\right)\boldsymbol{F}_{i-1}\right\|_{\infty} = \left\|\sum_{i=1}^{j_0}\frac{1}{q}\mathcal{A}_{\Omega_i^{\mathrm{clean}}}\boldsymbol{F}_{i-1}\right\|_{\infty} \\
&\leq \sum_{i=1}^{j_0}\frac{1}{q}\max_{(k,l)\in[n_1]\times[n_2]}\frac{\left|\left\langle\boldsymbol{A}_{(k,l)},\boldsymbol{F}_{i-1}\right\rangle\right|}{\sqrt{\omega_{k,l}}} = \sum_{i=1}^{j_0}\frac{1}{q}\sqrt{\nu\left(\boldsymbol{F}_{i-1}\right)} \\
&\leq \sum_{i=1}^{j_0}\frac{5\log\left(n_1n_2\right)}{\rho}\sqrt{\frac{\max\left(\mu_1c_{\mathrm{s}},\mu_2,\mu_4\right)r}{4^{i-1}n_1^2n_2^2}} \\
&\leq \sum_{i=0}^{j_0-1}\frac{1}{2^i}\sqrt{\frac{25\max\left(\mu_1c_{\mathrm{s}},\mu_2,\mu_4\right)r\log^2\left(n_1n_2\right)}{m^2}} \\
&< \frac{1}{4}\sqrt{\frac{1}{m\log\left(n_1n_2\right)}} = \frac{\lambda}{4},
\end{aligned}$$

when $m > c_{25}\max\left(\mu_1c_{\mathrm{s}},\mu_2,\mu_4\right)r\log^3\left(n_1n_2\right)$ for some constant $c_{25}>0$.

We have verified that $\boldsymbol{W}$ satisfies the four conditions required in (44), and is hence a valid dual certificate. This completes the proof.

8 Concluding Remarks

We present an efficient nonparametric algorithm to estimate a spectrally sparse object from its partial time-domain samples, which poses spectral compressed sensing as a low-rank Hankel structured matrix completion problem. Under mild incoherence conditions, our algorithm enables recovery of the multi-dimensional unknown frequencies with infinite precision, which remedies the basis mismatch issue that arises in conventional CS paradigms. We have shown both theoretically and numerically that our algorithm is stable against bounded noise and a constant portion of arbitrary corruptions, and can be extended to tasks such as super resolution. To the best of our knowledge, our result on Hankel matrix completion is also the first theoretical guarantee that is close to the information-theoretical limit (up to a logarithmic factor).

Our results are based on uniform random observation models. In particular, this paper considers directly taking a random subset of the time domain samples, it is also possible to take a random set of linear mixtures of the time domain samples, as in the renowned CS setting [7]. This again can be translated into taking linear measurements of the low-rank K-fold Hankel matrix, given as $\boldsymbol{y}=\mathcal{B}(\boldsymbol{X}_{\mathrm{e}})$. Unfortunately, due to the Hankel structures, $\mathcal{B}$ does not satisfy the matrix restricted isometry property (RIP) [34, 44]. Nonetheless, the technique developed in this paper can be extended without difficulty to analyze linear measurements, in a similar flavor of a golfing scheme developed for CS in [22].

It remains to be seen whether it is possible to obtain performance guarantees of the proposed EMaC algorithm similar to that in [24] for super resolution. It is also of great interest to develop efficient numerical methods to solve the EMaC algorithm in order to handle massive datasets.

A Proof of Lemma 2

(1) We first show that $\boldsymbol{E}_{\mathrm{L}}^*\boldsymbol{E}_{\mathrm{L}}$ and $\boldsymbol{E}_{\mathrm{R}}\boldsymbol{E}_{\mathrm{R}}^*$ coincide with the matrices $\boldsymbol{G}_{\mathrm{L}}$ and $\boldsymbol{G}_{R}^{T}$. Since $\boldsymbol{Y}_{\mathrm{d}}$ is a diagonal matrix, one can verify the identities

$$\left(\boldsymbol{Y}_{\mathrm{d}}^{l*}\boldsymbol{Z}_{\mathrm{L}}^*\boldsymbol{Z}_{\mathrm{L}}\boldsymbol{Y}_{\mathrm{d}}^{l}\right)_{i_1,i_2} = \left(y_{i_1}^*y_{i_2}\right)^l\left(\boldsymbol{Z}_{\mathrm{L}}^*\boldsymbol{Z}_{\mathrm{L}}\right)_{i_1,i_2},$$

and

$$\left(\boldsymbol{Z}_{\mathrm{L}}^*\boldsymbol{Z}_{\mathrm{L}}\right)_{i_1,i_2} = \sum_{k=0}^{k_2-1}\left(z_{i_1}^*z_{i_2}\right)^k = \begin{cases}\frac{1-\left(z_{i_1}^*z_{i_2}\right)^{k_2}}{1-z_{i_1}^*z_{i_2}}, & \text{if } i_1\neq i_2,\\ k_2, & \text{if } i_1=i_2,\end{cases}$$

which immediately give

$$\begin{aligned}
\boldsymbol{E}_{\mathrm{L}}^{*}\boldsymbol{E}_{\mathrm{L}} &= \frac{1}{k_1k_2}\left[\boldsymbol{Z}_{\mathrm{L}}^{*},\boldsymbol{Y}_{\mathrm{d}}^{*}\boldsymbol{Z}_{\mathrm{L}}^{*},\cdots,\left(\boldsymbol{Y}_{\mathrm{d}}^{*}\right)^{k_1-1}\boldsymbol{Z}_{\mathrm{L}}^{*}\right]\left[\begin{array}{c}\boldsymbol{Z}_L\\ \boldsymbol{Z}_L\boldsymbol{Y}_{\mathrm{d}}\\ \vdots\\ \boldsymbol{Z}_L\boldsymbol{Y}_{\mathrm{d}}^{k_1-1}\end{array}\right] = \frac{1}{k_1k_2}\sum_{l=0}^{k_1-1}\boldsymbol{Y}_{\mathrm{d}}^{l*}\boldsymbol{Z}_{\mathrm{L}}^{*}\boldsymbol{Z}_{\mathrm{L}}\boldsymbol{Y}_{\mathrm{d}}^{l}\\
&= \frac{1}{k_1k_2}\left(\left(\sum_{l=0}^{k_1-1}\left(y_{i_1}^{*}y_{i_2}\right)^{l}\right)\left(\boldsymbol{Z}_{\mathrm{L}}^{*}\boldsymbol{Z}_{\mathrm{L}}\right)_{i_1,i_2}\right)_{1\le i_1,i_2\le r}\\
&= \frac{1}{k_1k_2}\left(\frac{1-\left(y_{i_1}^{*}y_{i_2}\right)^{k_1}}{1-y_{i_1}^{*}y_{i_2}}\frac{1-\left(z_{i_1}^{*}z_{i_2}\right)^{k_2}}{1-z_{i_1}^{*}z_{i_2}}\right)_{1\le i_1,i_2\le r}
\end{aligned}$$

with the convention that $\frac{1-\left(y_{i_1}^{*}y_{i_1}\right)^{k_1}}{1-y_{i_1}^{*}y_{i_1}}=k_1$ and $\frac{1-\left(z_{i_1}^{*}z_{i_1}\right)^{k_2}}{1-z_{i_1}^{*}z_{i_1}}=k_2$. That said, all diagonal entries satisfy $\left(\boldsymbol{E}_{\mathrm{L}}^{*}\boldsymbol{E}_{\mathrm{L}}\right)_{i_1,i_1}=1$, and the magnitude of off-diagonal entries can be calculated as

$$\left|\left(\boldsymbol{E}_{\mathrm{L}}^{*}\boldsymbol{E}_{\mathrm{L}}\right)_{i_1,i_2}\right| = \left|\frac{\sin\left[\pi k_1\left(f_{1i_1}-f_{1i_2}\right)\right]}{k_1\sin\left[\pi\left(f_{1i_1}-f_{1i_2}\right)\right]}\frac{\sin\left[\pi k_2\left(f_{2i_1}-f_{2i_2}\right)\right]}{k_2\sin\left[\pi\left(f_{2i_1}-f_{2i_2}\right)\right]}\right|.$$

Recall that this exactly coincides with the definition of $\boldsymbol{G}_{\mathrm{L}}$. Similarly, $\boldsymbol{G}_{\mathrm{R}}=\left(\boldsymbol{E}_{\mathrm{R}}\boldsymbol{E}_{\mathrm{R}}^{*}\right)^{T}$. These findings immediately yield

$$\sigma_{\min}\left(\boldsymbol{E}_{\mathrm{L}}^{*}\boldsymbol{E}_{\mathrm{L}}\right)\ge\frac{1}{\mu_1},\quad\text{and}\quad\sigma_{\min}\left(\boldsymbol{E}_{\mathrm{R}}\boldsymbol{E}_{\mathrm{R}}^{*}\right)\ge\frac{1}{\mu_1}. \tag{51}$$

(2) Consider the case in which we only know $\sigma_{\min}\left(\boldsymbol{G}_{\mathrm{L}}\right)\ge\frac{1}{\mu_1}$ and $\sigma_{\min}\left(\boldsymbol{G}_{\mathrm{R}}\right)\ge\frac{1}{\mu_1}$. In fact, since $\left|\langle\boldsymbol{A}_{\boldsymbol{b}},\mathcal{P}_T\boldsymbol{A}_{\boldsymbol{a}}\rangle\right|=\left|\langle\mathcal{P}_T\boldsymbol{A}_{\boldsymbol{b}},\boldsymbol{A}_{\boldsymbol{a}}\rangle\right|$, we only need to examine the situation where $\omega_{\boldsymbol{b}}<\omega_{\boldsymbol{a}}$.

Observe that

$$\left|\langle\boldsymbol{A}_{\boldsymbol{b}},\mathcal{P}_T\boldsymbol{A}_{\boldsymbol{a}}\rangle\right|\le\left|\langle\boldsymbol{A}_{\boldsymbol{b}},\boldsymbol{U}\boldsymbol{U}^{*}\boldsymbol{A}_{\boldsymbol{a}}\rangle\right|+\left|\langle\boldsymbol{A}_{\boldsymbol{b}},\boldsymbol{A}_{\boldsymbol{a}}\boldsymbol{V}\boldsymbol{V}^{*}\rangle\right|+\left|\langle\boldsymbol{A}_{\boldsymbol{b}},\boldsymbol{U}\boldsymbol{U}^{*}\boldsymbol{A}_{\boldsymbol{a}}\boldsymbol{V}\boldsymbol{V}^{*}\rangle\right|.$$

Owing to the multi-fold Hankel structure of $\boldsymbol{A}_{\boldsymbol{a}}$, the matrix $\boldsymbol{U}\boldsymbol{U}^{*}\sqrt{\omega_{\boldsymbol{a}}}\boldsymbol{A}_{\boldsymbol{a}}$ consists of $\omega_{\boldsymbol{a}}$ columns of $\boldsymbol{U}\boldsymbol{U}^{*}$. Since there are only $\omega_{\boldsymbol{b}}$ nonzero entries in $\boldsymbol{A}_{\boldsymbol{b}}$ each of magnitude $\frac{1}{\sqrt{\omega_{\boldsymbol{b}}}}$, we can derive

$$\begin{aligned}
\left|\langle\boldsymbol{A}_{\boldsymbol{b}},\boldsymbol{U}\boldsymbol{U}^{*}\boldsymbol{A}_{\boldsymbol{a}}\rangle\right| &\le\left\|\boldsymbol{A}_{\boldsymbol{b}}\right\|_1\left\|\boldsymbol{U}\boldsymbol{U}^{*}\boldsymbol{A}_{\boldsymbol{a}}\right\|_{\infty}=\omega_{\boldsymbol{b}}\cdot\frac{1}{\sqrt{\omega_{\boldsymbol{b}}}}\cdot\max_{\alpha,\beta}\left|\left(\boldsymbol{U}\boldsymbol{U}^{*}\boldsymbol{A}_{\boldsymbol{a}}\right)_{\alpha,\beta}\right|\\
&\le\sqrt{\frac{\omega_{\boldsymbol{b}}}{\omega_{\boldsymbol{a}}}}\max_{\alpha,\beta}\left|\left(\boldsymbol{U}\boldsymbol{U}^{*}\right)_{\alpha,\beta}\right|.
\end{aligned}$$

Denote by $\boldsymbol{M}_{*k}$ and $\boldsymbol{M}_{k*}$ the kth column and kth row of $\boldsymbol{M}$, respectively, then it can be observed that each entry of $\boldsymbol{U}\boldsymbol{U}^{*}$ is bounded in magnitude by

$$\begin{aligned}
\left|\left(\boldsymbol{U}\boldsymbol{U}^{*}\right)_{k,l}\right| &= \left|\left(\boldsymbol{E}_{\mathrm{L}}\left(\boldsymbol{E}_{\mathrm{L}}^{*}\boldsymbol{E}_{\mathrm{L}}\right)^{-1}\boldsymbol{E}_{\mathrm{L}}^{*}\right)_{k,l}\right| = \left|\left(\boldsymbol{E}_{\mathrm{L}}\right)_{k*}\left(\boldsymbol{E}_{\mathrm{L}}^{*}\boldsymbol{E}_{\mathrm{L}}\right)^{-1}\left(\left(\boldsymbol{E}_{\mathrm{L}}\right)_{l*}\right)^{*}\right|\\
&\le\left\|\left(\boldsymbol{E}_{\mathrm{L}}\right)_{k*}\right\|_{\mathrm{F}}\left\|\left(\boldsymbol{E}_{\mathrm{L}}\right)_{l*}\right\|_{\mathrm{F}}\left\|\left(\boldsymbol{E}_{\mathrm{L}}^{*}\boldsymbol{E}_{\mathrm{L}}\right)^{-1}\right\|\\
&\le\frac{r}{k_1k_2}\frac{1}{\sigma_{\min}\left(\boldsymbol{E}_{\mathrm{L}}^{*}\boldsymbol{E}_{\mathrm{L}}\right)}\le\frac{\mu_1c_{\mathrm{s}}r}{n_1n_2},
\end{aligned} \tag{52}$$

which immediately implies that

$$\left|\langle\boldsymbol{A}_{\boldsymbol{b}},\boldsymbol{U}\boldsymbol{U}^{*}\boldsymbol{A}_{\boldsymbol{a}}\rangle\right|\le\sqrt{\frac{\omega_{\boldsymbol{b}}}{\omega_{\boldsymbol{a}}}}\frac{\mu_1c_{\mathrm{s}}r}{n_1n_2}. \tag{53}$$

Similarly, one can derive

$$\left|\langle\boldsymbol{A}_{\boldsymbol{b}},\boldsymbol{A}_{\boldsymbol{a}}\boldsymbol{V}\boldsymbol{V}^{*}\rangle\right|\le\sqrt{\frac{\omega_{\boldsymbol{b}}}{\omega_{\boldsymbol{a}}}}\frac{\mu_1c_{\mathrm{s}}r}{n_1n_2}. \tag{54}$$

We still need to bound the magnitude of $\langle \boldsymbol{U}\boldsymbol{U}^{*}\boldsymbol{A}_{\boldsymbol{a}}\boldsymbol{V}\boldsymbol{V}^{*},\boldsymbol{A}_{\boldsymbol{b}}\rangle$. One can observe that for any $1\leq k\leq k_1k_2$:

$$\begin{aligned}
\left\|(\boldsymbol{U}\boldsymbol{U}^{*})_{k*}\right\|_{\mathrm{F}} &\leq \left\|\left(\boldsymbol{E}_{\mathrm{L}}\right)_{k*}\left(\boldsymbol{E}_{\mathrm{L}}^{*}\boldsymbol{E}_{\mathrm{L}}\right)^{-1}\boldsymbol{E}_{\mathrm{L}}^{*}\right\|_{\mathrm{F}} \\
&\leq \left\|\left(\boldsymbol{E}_{\mathrm{L}}\right)_{k*}\right\|_{\mathrm{F}}\left\|\left(\boldsymbol{E}_{\mathrm{L}}^{*}\boldsymbol{E}_{\mathrm{L}}\right)^{-1}\boldsymbol{E}_{\mathrm{L}}^{*}\right\| \leq \sqrt{\frac{r}{k_1k_2}}\cdot\frac{1}{\sqrt{\sigma_{\min}\left(\boldsymbol{E}_{\mathrm{L}}^{*}\boldsymbol{E}_{\mathrm{L}}\right)}} \\
&\leq \sqrt{\frac{c_{\mathrm{s}}r}{n_1n_2\sigma_{\min}\left(\boldsymbol{E}_{\mathrm{L}}^{*}\boldsymbol{E}_{\mathrm{L}}\right)}}.
\end{aligned}$$

Similarly, for any $1\leq l\leq (n_1-k_1+1)(n_2-k_2+1)$, one has $\left\|(\boldsymbol{V}\boldsymbol{V}^{*})_{*l}\right\|_{\mathrm{F}}\leq\sqrt{\frac{c_{\mathrm{s}}r}{n_1n_2\sigma_{\min}\left(\boldsymbol{E}_{\mathrm{L}}^{*}\boldsymbol{E}_{\mathrm{L}}\right)}}$. The magnitude of all entries of $\boldsymbol{U}\boldsymbol{U}^{*}\boldsymbol{A}_{\boldsymbol{a}}\boldsymbol{V}\boldsymbol{V}^{*}$ can now be bounded by

$$\begin{aligned}
\max_{k,l}\left|\left(\boldsymbol{U}\boldsymbol{U}^{*}\boldsymbol{A}_{\boldsymbol{a}}\boldsymbol{V}\boldsymbol{V}^{*}\right)_{k,l}\right| &\leq \left\|\boldsymbol{A}_{\boldsymbol{a}}\right\|\max_{k}\left\|(\boldsymbol{U}\boldsymbol{U}^{*})_{k*}\right\|_{\mathrm{F}}\max_{l}\left\|(\boldsymbol{V}\boldsymbol{V}^{*})_{*l}\right\|_{\mathrm{F}} \\
&\leq \frac{1}{\sqrt{\omega_{\boldsymbol{a}}}}\frac{c_{\mathrm{s}}r}{n_1n_2\sigma_{\min}\left(\boldsymbol{E}_{\mathrm{L}}^{*}\boldsymbol{E}_{\mathrm{L}}\right)} \\
&\leq \frac{1}{\sqrt{\omega_{\boldsymbol{a}}}}\frac{\mu_1c_{\mathrm{s}}r}{n_1n_2}.
\end{aligned}$$

Since $\boldsymbol{A}_{\boldsymbol{b}}$ has only $\omega_{\boldsymbol{b}}$ nonzero entries each has magnitude $\frac{1}{\sqrt{\omega_{\boldsymbol{b}}}}$, one can verify that

$$\left|\langle \boldsymbol{U}\boldsymbol{U}^{*}\boldsymbol{A}_{\boldsymbol{a}}\boldsymbol{V}\boldsymbol{V}^{*},\boldsymbol{A}_{\boldsymbol{b}}\rangle\right| \leq \left(\max_{k,l}\left|\left(\boldsymbol{U}\boldsymbol{U}^{*}\boldsymbol{A}_{\boldsymbol{a}}\boldsymbol{V}\boldsymbol{V}^{*}\right)_{k,l}\right|\right)\cdot\frac{1}{\sqrt{\omega_{\boldsymbol{b}}}}\omega_{\boldsymbol{b}} = \sqrt{\frac{\omega_{\boldsymbol{b}}}{\omega_{\boldsymbol{a}}}}\frac{\mu_1c_{\mathrm{s}}r}{n_1n_2}. \tag{55}$$

The above bounds (53), (54) and (55) taken together lead to

$$\begin{aligned}
\left|\langle \boldsymbol{A}_{\boldsymbol{b}},\mathcal{P}_T\boldsymbol{A}_{\boldsymbol{a}}\rangle\right| &\leq \left|\langle \boldsymbol{U}\boldsymbol{U}^{*}\boldsymbol{A}_{\boldsymbol{a}},\boldsymbol{A}_{\boldsymbol{b}}\rangle\right| + \left|\langle \boldsymbol{A}_{\boldsymbol{a}}\boldsymbol{V}\boldsymbol{V}^{*},\boldsymbol{A}_{\boldsymbol{b}}\rangle\right| + \left|\langle \boldsymbol{U}\boldsymbol{U}^{*}\boldsymbol{A}_{\boldsymbol{a}}\boldsymbol{V}\boldsymbol{V}^{*},\boldsymbol{A}_{\boldsymbol{b}}\rangle\right| \\
&\leq \sqrt{\frac{\omega_{\boldsymbol{b}}}{\omega_{\boldsymbol{a}}}}\frac{3\mu_1c_{\mathrm{s}}r}{n_1n_2}.
\end{aligned} \tag{56}$$

(3) On the other hand, the bound on $\left|\langle \boldsymbol{A}_{\boldsymbol{b}},\mathcal{P}_T\boldsymbol{A}_{\boldsymbol{a}}\rangle\right|$ immediately leads the following upper bounds on $\sum_{\boldsymbol{a}}\left|\langle \boldsymbol{U}\boldsymbol{U}^{*}\boldsymbol{A}_{\boldsymbol{a}}\boldsymbol{V}\boldsymbol{V}^{*},\boldsymbol{A}_{\boldsymbol{b}}\rangle\right|^2\omega_{\boldsymbol{a}}$ and $\sum_{\boldsymbol{a}}\left|\langle \mathcal{P}_T\boldsymbol{A}_{\boldsymbol{b}},\boldsymbol{A}_{\boldsymbol{a}}\rangle\right|^2\omega_{\boldsymbol{a}}$:

$$\begin{aligned}
&\sum_{\boldsymbol{a}\in[n_1]\times[n_2]}\left|\langle \boldsymbol{U}\boldsymbol{U}^{*}\boldsymbol{A}_{\boldsymbol{a}}\boldsymbol{V}\boldsymbol{V}^{*},\boldsymbol{A}_{\boldsymbol{b}}\rangle\right|^2\omega_{\boldsymbol{a}} \\
\leq &\sum_{\boldsymbol{a}\in[n_1]\times[n_2]}\left(\sqrt{\frac{\omega_{\boldsymbol{b}}}{\omega_{\boldsymbol{a}}}}\frac{\mu_1c_{\mathrm{s}}r}{n_1n_2}\right)^2\omega_{\boldsymbol{a}} = \omega_{\boldsymbol{b}}\sum_{\boldsymbol{a}\in[n_1]\times[n_2]}\left(\frac{\mu_1c_{\mathrm{s}}r}{n_1n_2}\right)^2 \\
= &\omega_{\boldsymbol{b}}\frac{\mu_1^2c_{\mathrm{s}}^2r^2}{n_1n_2}
\end{aligned}$$

which simply come from the inequality (55), and

$$\begin{aligned}
&\sum_{\boldsymbol{a}\in[n_1]\times[n_2]}\left|\langle \mathcal{P}_T\boldsymbol{A}_{\boldsymbol{b}},\boldsymbol{A}_{\boldsymbol{a}}\rangle\right|^2\omega_{\boldsymbol{a}} \\
\leq &\sum_{\boldsymbol{a}\in[n_1]\times[n_2]}\left(\sqrt{\frac{\omega_{\boldsymbol{b}}}{\omega_{\boldsymbol{a}}}}\frac{3\mu_1c_{\mathrm{s}}r}{n_1n_2}\right)^2\omega_{\boldsymbol{a}} = \omega_{\boldsymbol{b}}\sum_{\boldsymbol{a}\in[n_1]\times[n_2]}\left(\frac{3\mu_1c_{\mathrm{s}}r}{n_1n_2}\right)^2 \\
= &\omega_{\boldsymbol{b}}\frac{9\mu_1^2c_{\mathrm{s}}^2r^2}{n_1n_2},
\end{aligned}$$

which is an immediate consequence of (56). These bounds indicate that $\mu_3\leq\mu_1^2c_{\mathrm{s}}^2r$ and $\mu_4\leq 9\mu_1^2c_{\mathrm{s}}^2r$.

We can also obtain an upper bound on μ_2 through μ_1 as follows. Observe that there exists a unitary matrix $\boldsymbol{B}$ such that

$$\boldsymbol{U}\boldsymbol{V}^{*} = \boldsymbol{E}_{\mathrm{L}}\left(\boldsymbol{E}_{\mathrm{L}}^{*}\boldsymbol{E}_{\mathrm{L}}\right)^{-\frac{1}{2}}\boldsymbol{B}\left(\boldsymbol{E}_{\mathrm{R}}\boldsymbol{E}_{\mathrm{R}}^{*}\right)^{-\frac{1}{2}}\boldsymbol{E}_{\mathrm{R}}.$$

For any $(k,l)\in[n_1]\times[n_2]$, we can then bound

$$\begin{aligned}
\left|(\boldsymbol{U}\boldsymbol{V}^*)_{k,l}\right| &= \left|\left(\boldsymbol{E}_{\mathrm{L}}\left(\boldsymbol{E}_{\mathrm{L}}^*\boldsymbol{E}_{\mathrm{L}}\right)^{-\frac{1}{2}}\boldsymbol{B}\left(\boldsymbol{E}_{\mathrm{R}}\boldsymbol{E}_{\mathrm{R}}^*\right)^{-\frac{1}{2}}\boldsymbol{E}_{\mathrm{R}}\right)_{k,l}\right| \\
&\le \|(\boldsymbol{E}_{\mathrm{L}})_{k*}\|_{\mathrm{F}}\left\|\left(\boldsymbol{E}_{\mathrm{L}}^*\boldsymbol{E}_{\mathrm{L}}\right)^{-\frac{1}{2}}\right\|\|\boldsymbol{B}\|\left\|\left(\boldsymbol{E}_{\mathrm{R}}^*\boldsymbol{E}_{\mathrm{R}}\right)^{-\frac{1}{2}}\right\|\|(\boldsymbol{E}_{\mathrm{R}})_{*l}\|_{\mathrm{F}} \\
&\le \sqrt{\frac{r}{k_1k_2}}\mu_1\sqrt{\frac{r}{(n_1-k_1+1)(n_2-k_2+1)}} \\
&\le \frac{\mu_1 c_{\mathrm{s}} r}{n_1 n_2}.
\end{aligned}$$

Since $\boldsymbol{A}_{(k,l)}$ has only $\omega_{k,l}$ nonzero entries each of magnitude $\frac{1}{\sqrt{\omega_{k,l}}}$, this leads to

$$\begin{aligned}
\frac{1}{\omega_{k,l}^2}\left|\sum_{(\alpha,\beta)\in\Omega_{\mathrm{e}}(k,l)}(\boldsymbol{U}\boldsymbol{V}^*)_{\alpha,\beta}\right|^2 &= \frac{1}{\omega_{k,l}}\left|\left\langle\boldsymbol{U}\boldsymbol{V}^*,\boldsymbol{A}_{(k,l)}\right\rangle\right|^2 \\
&\le \frac{1}{\omega_{k,l}}\left\{\left(\max_{k,l}\left|(\boldsymbol{U}\boldsymbol{V}^*)_{k,l}\right|\right)\frac{1}{\sqrt{\omega_{k,l}}}\cdot\omega_{k,l}\right\}^2 \\
&\le \left(\max_{k,l}\left|(\boldsymbol{U}\boldsymbol{V}^*)_{k,l}\right|\right)^2 \le \mu_1^2 c_{\mathrm{s}}^2 r\frac{r}{n_1^2n_2^2},
\end{aligned}$$

which indicates that $\mu_2\le\mu_1^2c_{\mathrm{s}}^2r$.

(4) Finally, we split $\sum_{\boldsymbol{a}\in[n_1]\times[n_2]}\left|\left\langle\mathcal{P}_T\boldsymbol{A}_{\boldsymbol{b}},\sqrt{\omega_{\boldsymbol{a}}}\boldsymbol{A}_{\boldsymbol{a}}\right\rangle\right|^2$ as follows

$$\begin{aligned}
&\sum_{\boldsymbol{a}\in[n_1]\times[n_2]}\left|\left\langle\mathcal{P}_T\boldsymbol{A}_{\boldsymbol{b}},\sqrt{\omega_{\boldsymbol{a}}}\boldsymbol{A}_{\boldsymbol{a}}\right\rangle\right|^2 = \sum_{\boldsymbol{a}\in[n_1]\times[n_2]}\left|\left\langle\left(\mathcal{P}_U+\mathcal{P}_V-\mathcal{P}_U\mathcal{P}_V\right)\boldsymbol{A}_{\boldsymbol{b}},\sqrt{\omega_{\boldsymbol{a}}}\boldsymbol{A}_{\boldsymbol{a}}\right\rangle\right|^2 \\
&\le 3\sum_{\boldsymbol{a}\in[n_1]\times[n_2]}\left\{\left|\left\langle\mathcal{P}_U\boldsymbol{A}_{\boldsymbol{b}},\sqrt{\omega_{\boldsymbol{a}}}\boldsymbol{A}_{\boldsymbol{a}}\right\rangle\right|^2+\left|\left\langle\mathcal{P}_V\boldsymbol{A}_{\boldsymbol{b}},\sqrt{\omega_{\boldsymbol{a}}}\boldsymbol{A}_{\boldsymbol{a}}\right\rangle\right|^2+\left|\left\langle\mathcal{P}_U\mathcal{P}_V\boldsymbol{A}_{\boldsymbol{b}},\sqrt{\omega_{\boldsymbol{a}}}\boldsymbol{A}_{\boldsymbol{a}}\right\rangle\right|^2\right\}
\end{aligned}$$

Now look at $\sum_{\boldsymbol{a}}\left|\left\langle\mathcal{P}_U\boldsymbol{A}_{\boldsymbol{b}},\sqrt{\omega_{\boldsymbol{a}}}\boldsymbol{A}_{\boldsymbol{a}}\right\rangle\right|^2=\sum_{\boldsymbol{a}}\left|\left\langle\boldsymbol{U}\boldsymbol{U}^*\boldsymbol{A}_{\boldsymbol{b}},\sqrt{\omega_{\boldsymbol{a}}}\boldsymbol{A}_{\boldsymbol{a}}\right\rangle\right|^2$. We know that

$$\|\boldsymbol{U}\boldsymbol{U}^*\boldsymbol{A}_{\boldsymbol{b}}\|_{\mathrm{F}}^2\le\frac{\mu_1c_{\mathrm{s}}r}{n_1n_2}, \tag{57}$$

and that $\boldsymbol{U}\boldsymbol{U}^*\boldsymbol{A}_{\boldsymbol{b}}$ has $\omega_{\boldsymbol{b}}$ non-zero columns, or,

$$\boldsymbol{U}\boldsymbol{U}^*\boldsymbol{A}_{\boldsymbol{b}}\overset{\text{column permutation}}{=}\frac{1}{\sqrt{\omega_{\boldsymbol{b}}}}\left[\underbrace{\overline{\boldsymbol{U}}_{\boldsymbol{b}}}_{\omega_{\boldsymbol{b}}\text{ columns}},\boldsymbol{0}\right], \tag{58}$$

and hence $\left\langle\boldsymbol{U}\boldsymbol{U}^*\boldsymbol{A}_{\boldsymbol{b}},\sqrt{\omega_{\boldsymbol{a}}}\boldsymbol{A}_{\boldsymbol{a}}\right\rangle$ is simply the sum of all entries of $\boldsymbol{U}\boldsymbol{U}^*\boldsymbol{A}_{\boldsymbol{b}}$ lying in the set $\Omega_{\mathrm{e}}(\boldsymbol{a})$. Since there are at most $\omega_{\boldsymbol{b}}$ nonzero entries (due to the above structure of $\boldsymbol{U}\boldsymbol{U}^*\boldsymbol{A}_{\boldsymbol{b}}$) in each sum, we can bound

$$\left|\left\langle\boldsymbol{U}\boldsymbol{U}^*\boldsymbol{A}_{\boldsymbol{b}},\sqrt{\omega_{\boldsymbol{a}}}\boldsymbol{A}_{\boldsymbol{a}}\right\rangle\right|^2=\left|\sum_{(\alpha,\beta)\in\Omega_{\mathrm{e}}(\boldsymbol{a})}(\boldsymbol{U}\boldsymbol{U}^*\boldsymbol{A}_{\boldsymbol{b}})_{\alpha,\beta}\right|^2\le\omega_{\boldsymbol{b}}\sum_{(\alpha,\beta)\in\Omega_{\mathrm{e}}(\boldsymbol{a})}\left|(\boldsymbol{U}\boldsymbol{U}^*\boldsymbol{A}_{\boldsymbol{b}})_{\alpha,\beta}\right|^2$$

using the inequality $\left(\sum_{i=1}^{\omega_{\boldsymbol{b}}}x_i\right)^2\le\omega_{\boldsymbol{b}}\sum_{i=1}^{\omega_{\boldsymbol{b}}}x_i^2$. This then gives

$$\begin{aligned}
\sum_{\boldsymbol{a}\in[n_1]\times[n_2]}\left|\left\langle\boldsymbol{U}\boldsymbol{U}^*\boldsymbol{A}_{\boldsymbol{b}},\sqrt{\omega_{\boldsymbol{a}}}\boldsymbol{A}_{\boldsymbol{a}}\right\rangle\right|^2 &\le \omega_{\boldsymbol{b}}\sum_{\boldsymbol{a}\in[n_1]\times[n_2]}\sum_{(\alpha,\beta)\in\Omega_{\mathrm{e}}(\boldsymbol{a})}\left|(\boldsymbol{U}\boldsymbol{U}^*\boldsymbol{A}_{\boldsymbol{b}})_{\alpha,\beta}\right|^2 \\
&\le \omega_{\boldsymbol{b}}\|\boldsymbol{U}\boldsymbol{U}^*\boldsymbol{A}_b\|_{\mathrm{F}}^2\le\omega_{\boldsymbol{b}}\frac{\mu_1c_{\mathrm{s}}r}{n_1n_2},
\end{aligned}$$

where the last inequality follows from Lemma 4. Similarly, one has

$$\sum_{\boldsymbol{a}\in[n_1]\times[n_2]}\left|\langle\boldsymbol{A_b}\boldsymbol{V}\boldsymbol{V}^*,\sqrt{\omega_{\boldsymbol{a}}}\boldsymbol{A_a}\rangle\right|^2\leq\omega_{\boldsymbol{b}}\frac{\mu_1c_{\mathrm{s}}r}{n_1n_2}.$$

To summarize,

$$\begin{aligned}&\sum_{\boldsymbol{a}\in[n_1]\times[n_2]}\left|\langle\mathcal{P}_T\boldsymbol{A_b},\sqrt{\omega_{\boldsymbol{a}}}\boldsymbol{A_a}\rangle\right|^2\\ \leq\ &3\sum_{\boldsymbol{a}\in[n_1]\times[n_2]}\left\{\left|\langle\mathcal{P}_U\boldsymbol{A_b},\sqrt{\omega_{\boldsymbol{a}}}\boldsymbol{A_a}\rangle\right|^2+\left|\langle\mathcal{P}_V\boldsymbol{A_b},\sqrt{\omega_{\boldsymbol{a}}}\boldsymbol{A_a}\rangle\right|^2+\left|\langle\mathcal{P}_U\mathcal{P}_V\boldsymbol{A_b},\sqrt{\omega_{\boldsymbol{a}}}\boldsymbol{A_a}\rangle\right|^2\right\}\\ \leq\ &\frac{6\mu_1c_{\mathrm{s}}\omega_{\boldsymbol{b}}r}{n_1n_2}+\frac{3\mu_3c_{\mathrm{s}}\omega_{\boldsymbol{b}}r}{n_1n_2}.\end{aligned}$$

B Proof of Lemma 3

Consider any valid perturbation $\boldsymbol{H}$ obeying $\mathcal{P}_{\Omega}(\boldsymbol{X}+\boldsymbol{H})=\mathcal{P}_{\Omega}(\boldsymbol{X})$, and denote by $\boldsymbol{H}_{\mathrm{e}}$ the enhanced form of $\boldsymbol{H}$. We note that the constraint requires $\mathcal{A}'_{\Omega}(\boldsymbol{H}_{\mathrm{e}})=0$ (or $\mathcal{A}_{\Omega}(\boldsymbol{H}_{\mathrm{e}})=0$) and $\mathcal{A}^{\perp}(\boldsymbol{H}_{\mathrm{e}})=0$. In addition, set $\boldsymbol{W}_0=\mathcal{P}_{T^{\perp}}(\boldsymbol{B})$ for any $\boldsymbol{B}$ that satisfies $\langle\boldsymbol{B},\mathcal{P}_{T^{\perp}}(\boldsymbol{H}_{\mathrm{e}})\rangle=\left\|\mathcal{P}_{T^{\perp}}(\boldsymbol{H}_{\mathrm{e}})\right\|_*$ and $\|\boldsymbol{B}\|\leq1$. Therefore, $\boldsymbol{W}_0\in T^{\perp}$ and $\|\boldsymbol{W}_0\|\leq1$, and hence $\boldsymbol{U}\boldsymbol{V}^*+\boldsymbol{W}_0$ is a subgradient of the nuclear norm at $\boldsymbol{X}_{\mathrm{e}}$. We will establish this lemma by considering two scenarios separately.

(1) Consider first the case in which $\boldsymbol{H}_{\mathrm{e}}$ satisfies

$$\left\|\mathcal{P}_T(\boldsymbol{H}_{\mathrm{e}})\right\|_{\mathrm{F}}\leq\frac{n_1^2n_2^2}{2}\left\|\mathcal{P}_{T^{\perp}}(\boldsymbol{H}_{\mathrm{e}})\right\|_{\mathrm{F}}.\tag{59}$$

Since $\boldsymbol{U}\boldsymbol{V}^*+\boldsymbol{W}_0$ is a subgradient of the nuclear norm at $\boldsymbol{X}_{\mathrm{e}}$, it follows that

$$\begin{aligned}\left\|\boldsymbol{X}_{\mathrm{e}}+\boldsymbol{H}_{\mathrm{e}}\right\|_*&\geq\left\|\boldsymbol{X}_{\mathrm{e}}\right\|_*+\langle\boldsymbol{U}\boldsymbol{V}^*+\boldsymbol{W}_0,\boldsymbol{H}_{\mathrm{e}}\rangle\\&=\left\|\boldsymbol{X}_{\mathrm{e}}\right\|_*+\langle\boldsymbol{U}\boldsymbol{V}^*+\boldsymbol{W},\boldsymbol{H}_{\mathrm{e}}\rangle+\langle\boldsymbol{W}_0,\boldsymbol{H}_{\mathrm{e}}\rangle-\langle\boldsymbol{W},\boldsymbol{H}_{\mathrm{e}}\rangle\\&=\left\|\boldsymbol{X}_{\mathrm{e}}\right\|_*+\left\langle\left(\mathcal{A}'_{\Omega}+\mathcal{A}^{\perp}\right)(\boldsymbol{U}\boldsymbol{V}^*+\boldsymbol{W}),\boldsymbol{H}_{\mathrm{e}}\right\rangle+\langle\boldsymbol{W}_0,\boldsymbol{H}_{\mathrm{e}}\rangle-\langle\boldsymbol{W},\boldsymbol{H}_{\mathrm{e}}\rangle&&(60)\\&\geq\left\|\boldsymbol{X}_{\mathrm{e}}\right\|_*+\left\|\mathcal{P}_{T^{\perp}}(\boldsymbol{H}_{\mathrm{e}})\right\|_*-\langle\boldsymbol{W},\boldsymbol{H}_{\mathrm{e}}\rangle&&(61)\end{aligned}$$

where (60) holds from (35), and (61) follows from the property of $\boldsymbol{W}_0$ and the fact that $\left(\mathcal{A}'_{\Omega}+\mathcal{A}^{\perp}\right)(\boldsymbol{H}_{\mathrm{e}})=0$. The last term of (61) can be bounded as

$$\begin{aligned}\langle\boldsymbol{W},\boldsymbol{H}_{\mathrm{e}}\rangle&=\langle\mathcal{P}_T(\boldsymbol{W}),\boldsymbol{H}_{\mathrm{e}}\rangle+\langle\mathcal{P}_{T^{\perp}}(\boldsymbol{W}),\boldsymbol{H}_{\mathrm{e}}\rangle\\&\leq\left\|\mathcal{P}_T(\boldsymbol{W})\right\|_{\mathrm{F}}\left\|\mathcal{P}_T(\boldsymbol{H}_{\mathrm{e}})\right\|_{\mathrm{F}}+\left\|\mathcal{P}_{T^{\perp}}(\boldsymbol{W})\right\|\left\|\mathcal{P}_{T^{\perp}}(\boldsymbol{H}_{\mathrm{e}})\right\|_*\\&\leq\frac{1}{2n_1^2n_2^2}\left\|\mathcal{P}_T(\boldsymbol{H}_{\mathrm{e}})\right\|_{\mathrm{F}}+\frac{1}{2}\left\|\mathcal{P}_{T^{\perp}}(\boldsymbol{H}_{\mathrm{e}})\right\|_*,\end{aligned}$$

where the last inequality follows from the assumptions (36) and (37). Plugging this into (61) yields

$$\begin{aligned}\left\|\boldsymbol{X}_{\mathrm{e}}+\boldsymbol{H}_{\mathrm{e}}\right\|_*&\geq\left\|\boldsymbol{X}_{\mathrm{e}}\right\|_*-\frac{1}{2n_1^2n_2^2}\left\|\mathcal{P}_T(\boldsymbol{H}_{\mathrm{e}})\right\|_{\mathrm{F}}+\frac{1}{2}\left\|\mathcal{P}_{T^{\perp}}(\boldsymbol{H}_{\mathrm{e}})\right\|_*\\&\geq\left\|\boldsymbol{X}_{\mathrm{e}}\right\|_*-\frac{1}{4}\left\|\mathcal{P}_{T^{\perp}}(\boldsymbol{H}_{\mathrm{e}})\right\|_{\mathrm{F}}+\frac{1}{2}\left\|\mathcal{P}_{T^{\perp}}(\boldsymbol{H}_{\mathrm{e}})\right\|_{\mathrm{F}}&&(62)\\&\geq\left\|\boldsymbol{X}_{\mathrm{e}}\right\|_*+\frac{1}{4}\left\|\mathcal{P}_{T^{\perp}}(\boldsymbol{H}_{\mathrm{e}})\right\|_{\mathrm{F}}\end{aligned}$$

where (62) follows from the inequality $\|\boldsymbol{M}\|_*\geq\|\boldsymbol{M}\|_{\mathrm{F}}$ and (59). Therefore, $\boldsymbol{X}_{\mathrm{e}}$ is the minimizer of EMaC.

We still need to prove the uniqueness of the minimizer. The inequality (62) implies that $\left\|\boldsymbol{X}_{\mathrm{e}}+\boldsymbol{H}_{\mathrm{e}}\right\|_*=\left\|\boldsymbol{X}_{\mathrm{e}}\right\|_*$ only when $\left\|\mathcal{P}_{T^{\perp}}(\boldsymbol{H}_{\mathrm{e}})\right\|_{\mathrm{F}}=0$. If $\left\|\mathcal{P}_{T^{\perp}}(\boldsymbol{H}_{\mathrm{e}})\right\|_{\mathrm{F}}=0$, then $\left\|\mathcal{P}_T(\boldsymbol{H}_{\mathrm{e}})\right\|_{\mathrm{F}}\leq\frac{n_1^2n_2^2}{2}\left\|\mathcal{P}_{T^{\perp}}(\boldsymbol{H}_{\mathrm{e}})\right\|_{\mathrm{F}}=0$, and

hence $\mathcal{P}_{T^{\perp}}\left(\boldsymbol{H}_{\mathrm{e}}\right)=\mathcal{P}_{T}\left(\boldsymbol{H}_{\mathrm{e}}\right)=0$, which only occurs when $\boldsymbol{H}_{\mathrm{e}}=0$. Hence, $\boldsymbol{X}_{\mathrm{e}}$ is the unique minimizer in this situation.

(2) On the other hand, consider the complement scenario where the following holds

$$\left\|\mathcal{P}_{T}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\|_{\mathrm{F}} \geq \frac{n_{1}^{2} n_{2}^{2}}{2}\left\|\mathcal{P}_{T^{\perp}}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\|_{\mathrm{F}}. \tag{63}$$

We would first like to bound $\left\|\left(\frac{n_{1} n_{2}}{m} \mathcal{A}_{\Omega}+\mathcal{A}^{\perp}\right) \mathcal{P}_{T}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\|_{\mathrm{F}}$ and $\left\|\left(\frac{n_{1} n_{2}}{m} \mathcal{A}_{\Omega}+\mathcal{A}^{\perp}\right) \mathcal{P}_{T^{\perp}}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\|_{\mathrm{F}}$. The former term can be lower bounded by

$$\begin{aligned}
& \left\|\left(\frac{n_{1} n_{2}}{m} \mathcal{A}_{\Omega}+\mathcal{A}^{\perp}\right) \mathcal{P}_{T}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\|_{\mathrm{F}}^{2} \\
= & \left\langle\left(\frac{n_{1} n_{2}}{m} \mathcal{A}_{\Omega}+\mathcal{A}^{\perp}\right) \mathcal{P}_{T}\left(\boldsymbol{H}_{\mathrm{e}}\right),\left(\frac{n_{1} n_{2}}{m} \mathcal{A}_{\Omega}+\mathcal{A}^{\perp}\right) \mathcal{P}_{T}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\rangle \\
= & \left\langle\frac{n_{1} n_{2}}{m} \mathcal{A}_{\Omega} \mathcal{P}_{T}\left(\boldsymbol{H}_{\mathrm{e}}\right), \frac{n_{1} n_{2}}{m} \mathcal{A}_{\Omega} \mathcal{P}_{T}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\rangle+\left\langle\mathcal{A}^{\perp} \mathcal{P}_{T}\left(\boldsymbol{H}_{\mathrm{e}}\right), \mathcal{A}^{\perp} \mathcal{P}_{T}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\rangle \\
\geq & \left\langle\mathcal{P}_{T}\left(\boldsymbol{H}_{\mathrm{e}}\right), \frac{n_{1} n_{2}}{m} \mathcal{A}_{\Omega} \mathcal{P}_{T}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\rangle+\left\langle\mathcal{P}_{T}\left(\boldsymbol{H}_{\mathrm{e}}\right), \mathcal{A}^{\perp} \mathcal{P}_{T}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\rangle \qquad (64)\\
= & \left\langle\mathcal{P}_{T}\left(\boldsymbol{H}_{\mathrm{e}}\right), \mathcal{P}_{T}\left(\frac{n_{1} n_{2}}{m} \mathcal{A}_{\Omega}+\mathcal{A}^{\perp}\right) \mathcal{P}_{T}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\rangle \\
= & \left\langle\mathcal{P}_{T}\left(\boldsymbol{H}_{\mathrm{e}}\right), \mathcal{P}_{T}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\rangle+\left\langle\mathcal{P}_{T}\left(\boldsymbol{H}_{\mathrm{e}}\right),\left(\frac{n_{1} n_{2}}{m} \mathcal{P}_{T} \mathcal{A}_{\Omega} \mathcal{P}_{T}-\mathcal{P}_{T} \mathcal{A} \mathcal{P}_{T}\right) \mathcal{P}_{T}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\rangle \\
\geq & \left\|\mathcal{P}_{T}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\|_{\mathrm{F}}^{2}-\left\|\mathcal{P}_{T} \mathcal{A} \mathcal{P}_{T}-\frac{n_{1} n_{2}}{m} \mathcal{P}_{T} \mathcal{A}_{\Omega} \mathcal{P}_{T}\right\|\left\|\mathcal{P}_{T}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\|_{\mathrm{F}}^{2} \qquad (65)\\
\geq & \left(1-\left\|\mathcal{P}_{T} \mathcal{A} \mathcal{P}_{T}-\frac{n_{1} n_{2}}{m} \mathcal{P}_{T} \mathcal{A}_{\Omega} \mathcal{P}_{T}\right\|\right)\left\|\mathcal{P}_{T}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\|_{\mathrm{F}}^{2} \\
\geq & \frac{1}{2}\left\|\mathcal{P}_{T}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\|_{\mathrm{F}}^{2}. \qquad (66)
\end{aligned}$$

On the other hand, since the operator norm of any projection operator is bounded above by 1, one can verify that

$$\left\|\frac{n_{1} n_{2}}{m} \mathcal{A}_{\Omega}+\mathcal{A}^{\perp}\right\| \leq \frac{n_{1} n_{2}}{m}\left(\left\|\mathcal{A}_{a_{1}}+\mathcal{A}^{\perp}\right\|+\sum_{i=2}^{m}\left\|\mathcal{A}_{a_{i}}\right\|\right) \leq n_{1} n_{2},$$

where a_{i} $(1 \leq i \leq m)$ are m uniform random indices that form Ω. This implies the following bound:

$$\left\|\left(\frac{n_{1} n_{2}}{m} \mathcal{A}_{\Omega}+\mathcal{A}^{\perp}\right) \mathcal{P}_{T^{\perp}}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\|_{\mathrm{F}} \leq n_{1} n_{2}\left\|\mathcal{P}_{T^{\perp}}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\|_{\mathrm{F}} \leq \frac{2}{n_{1} n_{2}}\left\|\mathcal{P}_{T}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\|_{\mathrm{F}}, \tag{67}$$

where the last inequality arises from our assumption. Combining this with the above two bounds yields

$$\begin{aligned}
0=\left\|\left(\frac{n_{1} n_{2}}{m} \mathcal{A}_{\Omega}+\mathcal{A}^{\perp}\right)\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\|_{\mathrm{F}} & \geq\left\|\left(\frac{n_{1} n_{2}}{m} \mathcal{A}_{\Omega}+\mathcal{A}^{\perp}\right) \mathcal{P}_{T}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\|_{\mathrm{F}}-\left\|\left(\frac{n_{1} n_{2}}{m} \mathcal{A}_{\Omega}+\mathcal{A}^{\perp}\right) \mathcal{P}_{T^{\perp}}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\|_{\mathrm{F}} \\
& \geq \sqrt{\frac{1}{2}}\left\|\mathcal{P}_{T}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\|_{\mathrm{F}}-\frac{2}{n_{1} n_{2}}\left\|\mathcal{P}_{T}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\|_{\mathrm{F}} \\
& \geq \frac{1}{2}\left\|\mathcal{P}_{T}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\|_{\mathrm{F}} \geq \frac{n_{1}^{2} n_{2}^{2}}{4}\left\|\mathcal{P}_{T^{\perp}}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\|_{\mathrm{F}} \geq 0,
\end{aligned}$$

which immediately indicates $\mathcal{P}_{T^{\perp}}\left(\boldsymbol{H}_{\mathrm{e}}\right)=0$ and $\mathcal{P}_{T}\left(\boldsymbol{H}_{\mathrm{e}}\right)=0$. Hence, (63) can only hold when $\boldsymbol{H}_{\mathrm{e}}=0$.

C Proof of Lemma 4

By definition, we have the identities

$$\begin{aligned}
\left\|\mathcal{P}_{T}\left(\boldsymbol{A}_{(k, l)}\right)\right\|_{\mathrm{F}}^{2} & =\left\langle\mathcal{P}_{T}\left(\boldsymbol{A}_{(k, l)}\right), \boldsymbol{A}_{(k, l)}\right\rangle \\
& =\left\langle\mathcal{P}_{U}\left(\boldsymbol{A}_{(k, l)}\right)+\mathcal{P}_{V}\left(\boldsymbol{A}_{(k, l)}\right)-\mathcal{P}_{U} \mathcal{P}_{V}\left(\boldsymbol{A}_{(k, l)}\right), \boldsymbol{A}_{(k, l)}\right\rangle \\
& =\left\|\mathcal{P}_{U}\left(\boldsymbol{A}_{(k, l)}\right)\right\|_{\mathrm{F}}^{2}+\left\|\mathcal{P}_{V}\left(\boldsymbol{A}_{(k, l)}\right)\right\|_{\mathrm{F}}^{2}-\left\|\mathcal{P}_{U} \mathcal{P}_{V}\left(\boldsymbol{A}_{(k, l)}\right)\right\|_{\mathrm{F}}^{2}.
\end{aligned}$$

Since $\boldsymbol{U}$ (resp. $\boldsymbol{V}$) and $\boldsymbol{E}_{\mathrm{L}}$ (resp. $\boldsymbol{E}_{\mathrm{R}}$) determine the same column (resp. row) space, we can write

$$\boldsymbol{U}\boldsymbol{U}^{*}=\boldsymbol{E}_{\mathrm{L}}\left(\boldsymbol{E}_{\mathrm{L}}^{*}\boldsymbol{E}_{\mathrm{L}}\right)^{-1}\boldsymbol{E}_{\mathrm{L}}^{*}\quad\text{and}\quad\boldsymbol{V}\boldsymbol{V}^{*}=\boldsymbol{E}_{\mathrm{R}}^{*}\left(\boldsymbol{E}_{\mathrm{R}}\boldsymbol{E}_{\mathrm{R}}^{*}\right)^{-1}\boldsymbol{E}_{\mathrm{R}},$$

and thus

$$\begin{aligned}\left\|\mathcal{P}_{T}\left(\boldsymbol{A}_{(k,l)}\right)\right\|_{\mathrm{F}}^{2}&\leq\left\|\mathcal{P}_{U}\left(\boldsymbol{A}_{(k,l)}\right)\right\|_{\mathrm{F}}^{2}+\left\|\mathcal{P}_{V}\left(\boldsymbol{A}_{(k,l)}\right)\right\|_{\mathrm{F}}^{2}\\&\leq\left\|\boldsymbol{E}_{\mathrm{L}}\left(\boldsymbol{E}_{\mathrm{L}}^{*}\boldsymbol{E}_{\mathrm{L}}\right)^{-1}\boldsymbol{E}_{\mathrm{L}}^{*}\boldsymbol{A}_{(k,l)}\right\|_{\mathrm{F}}^{2}+\left\|\boldsymbol{A}_{(k,l)}\boldsymbol{E}_{\mathrm{R}}^{*}\left(\boldsymbol{E}_{\mathrm{R}}\boldsymbol{E}_{\mathrm{R}}^{*}\right)^{-1}\boldsymbol{E}_{\mathrm{R}}\right\|_{\mathrm{F}}^{2}\\&\leq\frac{1}{\sigma_{\min}\left(\boldsymbol{E}_{\mathrm{L}}^{*}\boldsymbol{E}_{\mathrm{L}}\right)}\left\|\boldsymbol{E}_{\mathrm{L}}^{*}\boldsymbol{A}_{(k,l)}\right\|_{\mathrm{F}}^{2}+\frac{1}{\sigma_{\min}\left(\boldsymbol{E}_{\mathrm{R}}\boldsymbol{E}_{\mathrm{R}}^{*}\right)}\left\|\boldsymbol{A}_{(k,l)}\boldsymbol{E}_{\mathrm{R}}^{*}\right\|_{\mathrm{F}}^{2}.\end{aligned}$$

Note that $\sqrt{\omega_{k,l}}\boldsymbol{E}_{\mathrm{L}}^{*}\boldsymbol{A}_{(k,l)}$ consists of $\omega_{k,l}$ columns of $\boldsymbol{E}_{\mathrm{L}}^{*}$ (and hence it contains $r\omega_{k,l}$ nonzero entries in total). Owing to the fact that each entry of $\boldsymbol{E}_{\mathrm{L}}^{*}$ has magnitude $\frac{1}{\sqrt{k_{2}k_{2}}}$, one can derive

$$\left\|\boldsymbol{E}_{\mathrm{L}}^{*}\boldsymbol{A}_{(k,l)}\right\|_{\mathrm{F}}^{2}=\frac{1}{\omega_{k,l}}\cdot r\omega_{k,l}\cdot\frac{1}{k_{1}k_{2}}=\frac{r}{k_{1}k_{2}}\leq\frac{rc_{\mathrm{s}}}{n_{1}n_{2}}.$$

A similar argument yields

$$\left\|\boldsymbol{A}_{(k,l)}\boldsymbol{E}_{\mathrm{R}}^{*}\right\|_{\mathrm{F}}^{2}\leq\frac{c_{\mathrm{s}}r}{n_{1}n_{2}}.$$

We know from Lemma 2 that $\boldsymbol{E}_{\mathrm{L}}^{*}\boldsymbol{E}_{\mathrm{L}}=\boldsymbol{G}_{\mathrm{L}}$ and $\boldsymbol{E}_{\mathrm{R}}\boldsymbol{E}_{\mathrm{R}}^{*}=\boldsymbol{G}_{\mathrm{L}}^{T}$, and hence $\sigma_{\min}\left(\boldsymbol{E}_{\mathrm{L}}^{*}\boldsymbol{E}_{\mathrm{L}}\right)\geq\frac{1}{\mu_{1}}$ and $\sigma_{\min}\left(\boldsymbol{E}_{\mathrm{R}}\boldsymbol{E}_{\mathrm{R}}^{*}\right)\geq\frac{1}{\mu_{1}}$. One can, therefore, conclude that for every $(k,l)\in[n_{1}]\times[n_{2}]$,

$$\left\|\mathcal{P}_{U}\left(\boldsymbol{A}_{(k,l)}\right)\right\|_{\mathrm{F}}^{2}\leq\frac{\mu_{1}c_{\mathrm{s}}r}{n_{1}n_{2}},\quad\left\|\mathcal{P}_{V}\left(\boldsymbol{A}_{(k,l)}\right)\right\|_{\mathrm{F}}^{2}\leq\frac{\mu_{1}c_{\mathrm{s}}r}{n_{1}n_{2}},\quad\text{and}\quad\left\|\mathcal{P}_{T}\left(\boldsymbol{A}_{(k,l)}\right)\right\|_{\mathrm{F}}^{2}\leq\frac{2\mu_{1}c_{\mathrm{s}}r}{n_{1}n_{2}}.\tag{68}$$

D Proof of Lemma 5

Define a family of operators

$$\forall(k,l)\in[n_{1}]\times[n_{2}]:\quad\mathcal{Z}_{(k,l)}:=\frac{n_{1}n_{2}}{m}\mathcal{P}_{T}\mathcal{A}_{(k,l)}\mathcal{P}_{T}-\frac{1}{m}\mathcal{P}_{T}\mathcal{A}\mathcal{P}_{T}.$$

We can also compute

$$\mathcal{P}_{T}\mathcal{A}_{(k,l)}\mathcal{P}_{T}\left(\boldsymbol{M}\right)=\mathcal{P}_{T}\left\{\left\langle\boldsymbol{A}_{(k,l)},\mathcal{P}_{T}\boldsymbol{M}\right\rangle\boldsymbol{A}_{(k,l)}\right\}=\mathcal{P}_{T}\left(\boldsymbol{A}_{(k,l)}\right)\left\langle\mathcal{P}_{T}\left(\boldsymbol{A}_{(k,l)}\right),\boldsymbol{M}\right\rangle,\tag{69}$$

and hence

$$\begin{aligned}\left(\mathcal{P}_{T}\mathcal{A}_{(k,l)}\mathcal{P}_{T}\right)^{2}\left(\boldsymbol{M}\right)&=\left[\mathcal{P}_{T}\mathcal{A}_{(k,l)}\mathcal{P}_{T}\left\{\mathcal{P}_{T}\left(\boldsymbol{A}_{(k,l)}\right)\right\}\right]\left\langle\mathcal{P}_{T}\left(\boldsymbol{A}_{(k,l)}\right),\boldsymbol{M}\right\rangle\\&=\mathcal{P}_{T}\left\{\left\langle\boldsymbol{A}_{(k,l)},\mathcal{P}_{T}\left(\boldsymbol{A}_{(k,l)}\right)\right\rangle\boldsymbol{A}_{(k,l)}\right\}\left\langle\mathcal{P}_{T}\left(\boldsymbol{A}_{(k,l)}\right),\boldsymbol{M}\right\rangle\\&=\left\langle\boldsymbol{A}_{(k,l)},\mathcal{P}_{T}\left(\boldsymbol{A}_{(k,l)}\right)\right\rangle\mathcal{P}_{T}\left(\boldsymbol{A}_{(k,l)}\right)\left\langle\mathcal{P}_{T}\left(\boldsymbol{A}_{(k,l)}\right),\boldsymbol{M}\right\rangle.\end{aligned}\tag{70}$$

Comparing (69) and (70) gives

$$\left(\mathcal{P}_{T}\mathcal{A}_{(k,l)}\mathcal{P}_{T}\right)^{2}=\left\langle\boldsymbol{A}_{(k,l)},\mathcal{P}_{T}\left(\boldsymbol{A}_{(k,l)}\right)\right\rangle\mathcal{P}_{T}\mathcal{A}_{(k,l)}\mathcal{P}_{T}\leq\frac{2\mu_{1}c_{\mathrm{s}}r}{n_{1}n_{2}}\mathcal{P}_{T}\mathcal{A}_{(k,l)}\mathcal{P}_{T},\tag{71}$$

where the inequality follows from our assumption that

$$\left\langle\boldsymbol{A}_{(k,l)},\mathcal{P}_{T}\left(\boldsymbol{A}_{(k,l)}\right)\right\rangle=\left\|\mathcal{P}_{T}\left(\boldsymbol{A}_{(k,l)}\right)\right\|_{\mathrm{F}}^{2}\leq\frac{2\mu_{1}c_{\mathrm{s}}r}{n_{1}n_{2}}.$$

Let $\boldsymbol{a}_{i}$ ($1\leq i\leq m$) be m independent random pairs uniformly chosen from $[n_{1}]\times[n_{2}]$, then we have $\mathbb{E}\left(\mathcal{Z}_{\boldsymbol{a}_{i}}\right)=0$. This further gives

$$\begin{aligned}\mathbb{E}\mathcal{Z}_{\boldsymbol{a}_{i}}^{2}&=\mathbb{E}\left(\frac{n_{1}n_{2}}{m}\mathcal{P}_{T}\mathcal{A}_{\boldsymbol{a}_{i}}\mathcal{P}_{T}\right)^{2}-\left(\mathbb{E}\left(\frac{n_{1}n_{2}}{m}\mathcal{P}_{T}\mathcal{A}_{\boldsymbol{a}_{i}}\mathcal{P}_{T}\right)\right)^{2}\\&=\frac{n_{1}^{2}n_{2}^{2}}{m^{2}}\mathbb{E}\left(\mathcal{P}_{T}\mathcal{A}_{\boldsymbol{a}_{i}}\mathcal{P}_{T}\right)^{2}-\frac{1}{m^{2}}\left(\mathcal{P}_{T}\mathcal{A}\mathcal{P}_{T}\right)^{2},\end{aligned}$$

We can then bound the operator norm as

$$\begin{aligned}
\left\|\mathbb{E}\left(\mathcal{Z}_{\boldsymbol{a}_i}^2\right)\right\| &\leq \left\|\frac{n_1^2n_2^2}{m^2}\mathbb{E}\left(\mathcal{P}_T\mathcal{A}_{\boldsymbol{a}_i}\mathcal{P}_T\right)^2\right\| + \frac{1}{m^2}\left\|\left(\mathcal{P}_T\mathcal{A}\mathcal{P}_T\right)^2\right\| \\
&\leq \frac{n_1^2n_2^2}{m^2}\left\|\mathbb{E}\left(\mathcal{P}_T\mathcal{A}_{\boldsymbol{a}_i}\mathcal{P}_T\right)^2\right\| + \frac{1}{m^2} \\
&\leq \frac{n_1^2n_2^2}{m^2}\frac{2\mu_1c_{\mathrm{s}}r}{n_1n_2}\left\|\mathbb{E}\left(\mathcal{P}_T\mathcal{A}_{\boldsymbol{a}_i}\mathcal{P}_T\right)\right\| + \frac{1}{m^2} \qquad (72) \\
&= \frac{2\mu_1c_{\mathrm{s}}rn_1n_2}{m^2}\frac{1}{n_1n_2}\left\|\mathcal{P}_T\mathcal{A}\mathcal{P}_T\right\| + \frac{1}{m^2} \\
&\leq \frac{4\mu_1c_{\mathrm{s}}r}{m^2} := V_0, \qquad (73)
\end{aligned}$$

where (72) uses the fact that $\mathcal{P}_T\mathcal{A}_{\boldsymbol{a}_i}\mathcal{P}_T \succeq 0$. Besides, the first equality of (71) gives $\left\|\mathcal{P}_T\mathcal{A}_{(k,l)}\mathcal{P}_T\right\|^2 \leq \left\|\mathcal{P}_T\boldsymbol{A}_{(k,l)}\right\|_{\mathrm{F}}^2\left\|\mathcal{P}_T\mathcal{A}_{(k,l)}\mathcal{P}_T\right\|$ and hence $\left\|\mathcal{P}_T\mathcal{A}_{(k,l)}\mathcal{P}_T\right\| \leq \left\|\mathcal{P}_T\boldsymbol{A}_{(k,l)}\right\|_{\mathrm{F}}^2$, which immediately yields

$$\left\|\mathcal{Z}_{\boldsymbol{a}_i}\right\| \leq \frac{n_1n_2}{m}\left\|\mathcal{P}_T\boldsymbol{A}_{\boldsymbol{a}_i}\mathcal{P}_T\right\| + \frac{1}{m}\left\|\mathcal{P}_T\mathcal{A}\mathcal{P}_T\right\| \leq \frac{n_1n_2}{m}\left\|\mathcal{P}_T\boldsymbol{A}_{\boldsymbol{a}_i}\right\|_{\mathrm{F}}^2 + \frac{1}{m} < \frac{4\mu_1c_{\mathrm{s}}r}{m}.$$

This together with (73) gives

$$\frac{2mV_0}{\left\|\mathcal{Z}_{\boldsymbol{a}_i}\right\|} \geq 2.$$

Applying the Operator Bernstein Inequality [28, Theorem 6] yields that for any $t \leq 2$, we have

$$\mathbb{P}\left(\left\|\sum_{i=1}^{m}\mathcal{Z}_{\boldsymbol{a}_i}\right\| > t\right) \leq 2n_1n_2\exp\left(-\frac{t^2}{16\frac{\mu_1c_{\mathrm{s}}r}{m}}\right).$$

Finally, one can observe that $\sum_{i=1}^{m}\mathcal{Z}_{\boldsymbol{a}_i}$ is equivalent to $\frac{n_1n_2}{m}\mathcal{P}_T\mathcal{A}_\Omega\mathcal{P}_T - \mathcal{P}_T\mathcal{A}\mathcal{P}_T$ in distribution, which completes the proof.

E Proof of Lemma 6

Fix any $\boldsymbol{b} \in [n_1]\times[n_2]$. For any $\boldsymbol{a} \in [n_1]\times[n_2]$, define

$$z_{\boldsymbol{a}} = \frac{1}{qn_1n_2}\left\langle\boldsymbol{A}_b, \mathcal{P}_T\mathcal{A}\boldsymbol{F}\right\rangle - \left\langle\boldsymbol{A}_b, \frac{1}{q}\mathcal{P}_T\boldsymbol{A}_{\boldsymbol{a}}\right\rangle\left\langle\boldsymbol{A}_{\boldsymbol{a}}, \boldsymbol{F}\right\rangle.$$

Then for any i.i.d. α_i's chosen uniformly at random from $[n_1]\times[n_2]$, we can easily check that $\mathbb{E}\left(z_{\alpha_i}\right) = 0$. Define a multiset $\Omega_l := \{\alpha_i \mid 1 \leq i \leq qn_1n_2\}$, then the decomposition

$$\mathcal{A}_{\Omega_l}\boldsymbol{F} = \sum_{i=1}^{qn_1n_2}\boldsymbol{A}_{\alpha_i}\left\langle\boldsymbol{A}_{\alpha_i}, \boldsymbol{F}\right\rangle$$

allows us to derive

$$\left\langle\boldsymbol{A}_{\boldsymbol{b}}, \mathcal{P}_T\mathcal{A}_{\Omega_l}\boldsymbol{F}\right\rangle = \left\langle\boldsymbol{A}_{\boldsymbol{b}}, \sum_{i=1}^{qn_1n_2}\mathcal{P}_T\boldsymbol{A}_{\alpha_i}\left\langle\boldsymbol{A}_{\alpha_i}, \boldsymbol{F}\right\rangle\right\rangle,$$

and thus

$$\begin{aligned}
\sum_{i=1}^{qn_1n_2} z_{\alpha_i} &= \left\langle\boldsymbol{A}_{\boldsymbol{b}}, \mathcal{P}_T\mathcal{A}\boldsymbol{F}\right\rangle - \sum_{i=1}^{qn_1n_2}\left\langle\boldsymbol{A}_{\boldsymbol{b}}, \frac{1}{q}\mathcal{P}_T\boldsymbol{A}_{\alpha_i}\right\rangle\left\langle\boldsymbol{A}_{\alpha_i}, \boldsymbol{F}\right\rangle \\
&= \left\langle\boldsymbol{A}_{\boldsymbol{b}}, \mathcal{P}_T\mathcal{A}\boldsymbol{F}\right\rangle - \frac{1}{q}\left\langle\boldsymbol{A}_{\boldsymbol{b}}, \mathcal{P}_T\mathcal{A}_{\Omega_l}\boldsymbol{F}\right\rangle \\
&= \left\langle\boldsymbol{A}_{\boldsymbol{b}}, \left(\mathcal{P}_T\mathcal{A}\mathcal{P}_T - \frac{1}{q}\mathcal{P}_T\mathcal{A}_{\Omega_l}\mathcal{P}_T\right)\boldsymbol{F}\right\rangle.
\end{aligned}$$

Owing to the fact that $\mathbb{E}z_{\alpha_i} = 0$, we can bound the variance of each term as follows

$$\begin{aligned}\mathbb{E}\left|z_{\alpha_i}\right|^2 &= \mathbf{Var}\left(\left\langle \boldsymbol{A}_{\boldsymbol{b}}, \frac{1}{q}\mathcal{P}_T\boldsymbol{A}_{\alpha_i}\right\rangle \left\langle \boldsymbol{A}_{\alpha_i}, \boldsymbol{F}\right\rangle\right) \\ &\leq \mathbb{E}\left|\left\langle \boldsymbol{A}_{\boldsymbol{b}}, \frac{1}{q}\mathcal{P}_T\boldsymbol{A}_{\alpha_i}\right\rangle \left\langle \boldsymbol{A}_{\alpha_i}, \boldsymbol{F}\right\rangle\right|^2 \\ &= \frac{1}{n_1n_2}\sum_{\boldsymbol{a}\in[n_1]\times[n_2]}\left|\left\langle \boldsymbol{A}_{\boldsymbol{b}}, \frac{1}{q}\mathcal{P}_T\boldsymbol{A}_{\boldsymbol{a}}\right\rangle \left\langle \boldsymbol{A}_{\boldsymbol{a}}, \boldsymbol{F}\right\rangle\right|^2 \\ &\leq \frac{1}{q^2}\frac{\nu\left(\boldsymbol{F}\right)}{n_1n_2}\sum_{\boldsymbol{a}\in[n_1]\times[n_2]}\left|\left\langle \mathcal{P}_T\boldsymbol{A}_{\boldsymbol{b}}, \boldsymbol{A}_{\boldsymbol{a}}\right\rangle\right|^2\omega_{\boldsymbol{a}} \\ &\leq \frac{\mu_4 r\nu(\boldsymbol{F})}{\left(qn_1n_2\right)^2}\omega_{\boldsymbol{b}},\end{aligned}$$

where the last inequality arises from the definition of μ_4, i.e. for every $\boldsymbol{b}\in[n_1]\times[n_2]$,

$$\sum_{\boldsymbol{a}\in[n_1]\times[n_2]}\left|\left\langle \mathcal{P}_T\boldsymbol{A}_{\boldsymbol{b}}, \boldsymbol{A}_{\boldsymbol{a}}\right\rangle\right|^2\omega_{\boldsymbol{a}} \leq \frac{\mu_4 r}{n_1n_2}\omega_{\boldsymbol{b}}. \tag{74}$$

This immediately gives

$$\frac{1}{\omega_{\boldsymbol{b}}}\mathbb{E}\left(\sum_{i=1}^{qn_1n_2}\left|z_{\alpha_i}\right|^2\right) \leq \frac{\mu_4 r\nu(\boldsymbol{F})}{qn_1n_2} \leq \frac{\max\left\{\mu_4, 3\mu_1c_{\mathrm{s}}\right\}r\nu(\boldsymbol{F})}{qn_1n_2} := V.$$

On the other hand, Lemma 2 shows the inequality

$$\left|\left\langle \boldsymbol{A}_{\boldsymbol{b}}, \mathcal{P}_T\boldsymbol{A}_{\boldsymbol{a}}\right\rangle\right| \leq \sqrt{\frac{\omega_{\boldsymbol{b}}}{\omega_{\boldsymbol{a}}}\frac{3\mu_1c_{\mathrm{s}}r}{n_1n_2}}, \tag{75}$$

which further leads to

$$\begin{aligned}\frac{1}{\sqrt{\omega_{\boldsymbol{b}}}}\left|\left\langle \boldsymbol{A}_{\boldsymbol{b}}, \frac{1}{q}\mathcal{P}_T\boldsymbol{A}_{\boldsymbol{a}}\right\rangle \left\langle \boldsymbol{A}_{\boldsymbol{a}}, \boldsymbol{F}\right\rangle\right| &\leq \sqrt{\omega_{\boldsymbol{a}}\nu\left(\boldsymbol{F}\right)}\frac{1}{\sqrt{\omega_{\boldsymbol{b}}}q}\left|\left\langle \boldsymbol{A}_b, \mathcal{P}_T\boldsymbol{A}_{\boldsymbol{a}}\right\rangle\right| \\ &\leq \sqrt{\nu\left(\boldsymbol{F}\right)}\frac{1}{q}\frac{3\mu_1c_{\mathrm{s}}r}{n_1n_2}.\end{aligned}$$

Since $\frac{1}{qn_1n_2}\left\langle \boldsymbol{A}_{\boldsymbol{b}}, \mathcal{P}_T\mathcal{A}\boldsymbol{F}\right\rangle = \mathbb{E}\left\langle \boldsymbol{A}_{\boldsymbol{b}}, \frac{1}{q}\mathcal{P}_T\boldsymbol{A}_{\alpha_i}\right\rangle\left\langle \boldsymbol{A}_{\alpha_i}, \boldsymbol{F}\right\rangle$, one has as well

$$\frac{1}{\sqrt{\omega_{\boldsymbol{b}}}}\left|\frac{1}{qn_1n_2}\left\langle \boldsymbol{A}_{\boldsymbol{b}}, \mathcal{P}_T\mathcal{A}\boldsymbol{F}\right\rangle\right| = \frac{1}{\sqrt{\omega_{\boldsymbol{b}}}}\left|\mathbb{E}\left\langle \boldsymbol{A}_{\boldsymbol{b}}, \frac{1}{q}\mathcal{P}_T\boldsymbol{A}_{\boldsymbol{a}}\right\rangle \left\langle \boldsymbol{A}_{\boldsymbol{a}}, \boldsymbol{F}\right\rangle\right| \leq \sqrt{\nu\left(\boldsymbol{F}\right)}\frac{1}{q}\frac{3\mu_1c_{\mathrm{s}}r}{n_1n_2},$$

which immediately leads to

$$\begin{aligned}\frac{1}{\sqrt{\omega_{\boldsymbol{b}}}}\left|z_{\alpha_i}\right| &\leq \frac{1}{\sqrt{\omega_{\boldsymbol{b}}}}\left|\frac{1}{qn_1n_2}\left\langle \boldsymbol{A}_{\boldsymbol{b}}, \mathcal{P}_T\mathcal{A}\boldsymbol{F}\right\rangle\right| + \frac{1}{\sqrt{\omega_b}}\left|\left\langle \boldsymbol{A}_{\boldsymbol{b}}, \frac{1}{q}\mathcal{P}_T\boldsymbol{A}_{\alpha_i}\right\rangle \left\langle \boldsymbol{A}_{\alpha_i}, \boldsymbol{F}\right\rangle\right| \\ &\leq \sqrt{\nu\left(\boldsymbol{F}\right)}\frac{1}{q}\frac{6\mu_1c_{\mathrm{s}}r}{n_1n_2}\end{aligned}$$

The above bounds indicate that

$$\frac{2V}{\frac{1}{\sqrt{\omega_{\boldsymbol{b}}}}\left|z_{\boldsymbol{a}}\right|} \geq \sqrt{\nu(\boldsymbol{F})}.$$

Applying the operator Bernstein inequality [28, Theorem 6] yields for any $t < \nu\left(\boldsymbol{F}\right)$,

$$\mathbb{P}\left(\frac{1}{\omega_{\boldsymbol{b}}}\left|\sum_{i=1}^{qn_1n_2}z_{\alpha_i}\right|^2 > t\right) \leq c_6\exp\left(-\frac{tqn_1n_2}{4\max\left\{\mu_4, 3\mu_1c_{\mathrm{s}}\right\}r\nu(\boldsymbol{F})}\right).$$

Thus, there are some constants $c_7, \tilde{c}_7 > 0$ such that whenever $qn_1n_2 > \tilde{c}_7 \max\{\mu_4, 3\mu_1 c_\mathrm{s}\} r \log(n_1n_2)$ or, equivalently, $m > c_7 \max\{\mu_4, 3\mu_1 c_\mathrm{s}\} r \log^2(n_1n_2)$, we have

$$\mathbb{P}\left(\frac{\left|\sum_{i=1}^{qn_1n_2} z_{\alpha_i}\right|^2}{\omega_{\boldsymbol{b}}} > \frac{1}{4}\nu(\boldsymbol{F})\right) \leq \tilde{c}_6 \exp\left(-\frac{qn_1n_2}{16\max\{\mu_4, 3\mu_1 c_\mathrm{s}\} r\nu(\boldsymbol{F})}\right) \leq \frac{1}{(n_1n_2)^4}.$$

Finally, we observe that in distribution,

$$v\left(\left(\mathcal{P}_T\mathcal{A}\mathcal{P}_T - \frac{1}{q}\mathcal{P}_T\mathcal{A}_{\Omega_l}\mathcal{P}_T\right)\boldsymbol{F}\right) = \max_{\boldsymbol{b}\in[n_1]\times[n_2]} \frac{\left|\sum_{i=1}^{qn_1n_2} z_{\alpha_i}\right|^2}{\omega_{\boldsymbol{b}}}.$$

Applying a simple union bound over all $\boldsymbol{b} \in [n_1] \times [n_2]$ allows us to derive (41).

F Proof of Lemma 7

For any $\boldsymbol{a} \in [n_1] \times [n_2]$, define

$$\boldsymbol{H}_{\boldsymbol{a}} = \frac{1}{q}\mathcal{P}_{T^\perp}(\boldsymbol{A}_{\boldsymbol{a}})\langle \boldsymbol{A}_{\boldsymbol{a}}, \boldsymbol{F}\rangle + \frac{1}{qn_1n_2}\mathcal{P}_{T^\perp}\mathcal{A}^\perp(\boldsymbol{F}).$$

Let α_i ($1 \leq i \leq qn_1n_2$) be independently and uniformly drawn from $[n_1]\times[n_2]$ which forms Ω_l. Observing that

$$\mathcal{A}\boldsymbol{F} = \sum_{\boldsymbol{a}\in[n_1]\times[n_2]} \boldsymbol{A}_{\boldsymbol{a}}\langle \boldsymbol{A}_{\boldsymbol{a}}, \boldsymbol{F}\rangle,$$

we can write

$$\mathcal{P}_{T^\perp}\mathcal{A}\boldsymbol{F} = \sum_{\boldsymbol{a}\in[n_1]\times[n_2]} \mathcal{P}_{T^\perp}(\boldsymbol{A}_{\boldsymbol{a}})\langle \boldsymbol{A}_{\boldsymbol{a}}, \boldsymbol{F}\rangle.$$

This immediately gives

$$\begin{aligned}
\mathbb{E}\boldsymbol{H}_{\alpha_i} &= \frac{1}{qn_1n_2}\mathcal{P}_{T^\perp}\mathcal{A}^\perp(\boldsymbol{F}) + \frac{1}{qn_1n_2}\sum_{\boldsymbol{a}\in[n_1]\times[n_2]} \mathcal{P}_{T^\perp}(\boldsymbol{A}_{\boldsymbol{a}})\langle \boldsymbol{A}_{\boldsymbol{a}}, \boldsymbol{F}\rangle \\
&= \frac{1}{qn_1n_2}\mathcal{P}_{T^\perp}\mathcal{A}^\perp(\boldsymbol{F}) + \frac{1}{qn_1n_2}\mathcal{P}_{T^\perp}\mathcal{A}(\boldsymbol{F}) \\
&= \frac{1}{qn_1n_2}\mathcal{P}_{T^\perp}(\boldsymbol{F}) = 0.
\end{aligned}$$

Moreover, we have, in distribution, the following identity

$$\mathcal{P}_{T^\perp}\left(\frac{1}{q}\mathcal{A}_{\Omega_l} + \mathcal{A}^\perp\right)(\boldsymbol{F}) = \sum_{i=1}^{qn_1n_2} \boldsymbol{H}_{\alpha_i}.$$

On the other hand, since $\mathbb{E}\boldsymbol{H}_{\alpha_i} = 0$, if we denote $\mathcal{Y}_i = \frac{1}{q}\mathcal{P}_{T^\perp}(\boldsymbol{A}_{\alpha_i})\langle \boldsymbol{A}_{\alpha_i}, \boldsymbol{F}\rangle$, then $\boldsymbol{H}_{\alpha_i} = \mathcal{Y}_i - \mathbb{E}\mathcal{Y}_i$, and hence

$$\mathbb{E}\boldsymbol{H}_{\alpha_i}\boldsymbol{H}_{\alpha_i}^* = \mathbb{E}\left\{(\mathcal{Y}_i - \mathbb{E}\mathcal{Y}_i)(\mathcal{Y}_i - \mathbb{E}\mathcal{Y}_i)^*\right\} \leq \mathbb{E}\mathcal{Y}_i\mathcal{Y}_i^* = \frac{1}{q^2n_1n_2}\sum_{\boldsymbol{a}\in[n_1]\times[n_2]} |\langle \boldsymbol{A}_{\boldsymbol{a}}, \boldsymbol{F}\rangle|^2 \mathcal{P}_{T^\perp}(\boldsymbol{A}_{\boldsymbol{a}})(\mathcal{P}_{T^\perp}(\boldsymbol{A}_{\boldsymbol{a}}))^*.$$

The definition of the spectral norm $\|\boldsymbol{M}\| := \max_{\psi:\|\psi\|_2=1} \langle \psi, \boldsymbol{M}\psi \rangle$ allows us to bound

$$\begin{aligned}
\left\|\mathbb{E}\left(\boldsymbol{H}_{\alpha_i}\boldsymbol{H}_{\alpha_i}^*\right)\right\| &\leq \frac{1}{q^2}\max_{\psi:\|\psi\|_2=1}\left\{\frac{1}{n_1n_2}\sum_{\boldsymbol{a}\in[n_1]\times[n_2]}\left|\langle\boldsymbol{A}_{\boldsymbol{a}},\boldsymbol{F}\rangle\right|^2\left\langle\psi,\mathcal{P}_{T^\perp}\left(\boldsymbol{A}_{\boldsymbol{a}}\right)\left(\mathcal{P}_{T^\perp}\left(\boldsymbol{A}_{\boldsymbol{a}}\right)\right)^*\psi\right\rangle\right\} \\
&\leq \frac{1}{q^2n_1n_2}\nu(\boldsymbol{F})\max_{\psi:\|\psi\|_2=1}\left\langle\psi,\left(\sum_{\boldsymbol{a}\in[n_1]\times[n_2]}\omega_{\boldsymbol{a}}\mathcal{P}_{T^\perp}\left(\boldsymbol{A}_{\boldsymbol{a}}\right)\left(\mathcal{P}_{T^\perp}\left(\boldsymbol{A}_{\boldsymbol{a}}\right)\right)^*\right)\psi\right\rangle \\
&\leq \frac{1}{q^2n_1n_2}\nu(\boldsymbol{F})\left(\sum_{\boldsymbol{a}\in[n_1]\times[n_2]}\omega_{\boldsymbol{a}}\left\|\boldsymbol{A}_{\boldsymbol{a}}\right\|^2\right)\max_{\psi:\|\psi\|_2=1}\langle\psi,\psi\rangle \\
&\leq \frac{\nu(\boldsymbol{F})}{q^2},
\end{aligned}$$

where the last inequality uses the fact that $\|\boldsymbol{A}_{\boldsymbol{a}}\|^2 = \frac{1}{\omega_{\boldsymbol{a}}}$. Therefore,

$$\left\|\mathbb{E}\left(\sum_{i=1}^{qn_1n_2}\boldsymbol{H}_{\alpha_i}\boldsymbol{H}_{\alpha_i}^*\right)\right\| \leq \nu(\boldsymbol{F})n_1n_2\frac{1}{q} := V.$$

.

Besides, the definition (40) of $\nu(\boldsymbol{F})$ allows us to bound

$$\left\|\frac{1}{q}\mathcal{P}_{T^\perp}\left(\boldsymbol{A}_{\boldsymbol{a}}\right)\langle\boldsymbol{A}_{\boldsymbol{a}},\boldsymbol{F}\rangle\right\| \leq \sqrt{\nu\left(\boldsymbol{F}\right)\omega_{\boldsymbol{a}}}\frac{1}{q}\left\|\boldsymbol{A}_{\boldsymbol{a}}\right\| = \sqrt{\nu\left(\boldsymbol{F}\right)}\frac{1}{q}.$$

The fact that $\mathbb{E}\boldsymbol{H}_{\alpha_i} = 0$ yields

$$\left\|\frac{1}{qn_1n_2}\mathcal{P}_{T^\perp}\mathcal{A}^\perp\left(\boldsymbol{F}\right)\right\| = \left\|\mathbb{E}\frac{1}{q}\mathcal{P}_{T^\perp}\left(\boldsymbol{A}_{\alpha_i}\right)\langle\boldsymbol{A}_{\alpha_i},\boldsymbol{F}\rangle\right\| \leq \sqrt{\nu\left(\boldsymbol{F}\right)}\frac{1}{q},$$

and hence

$$\|\boldsymbol{H}_{\alpha_i}\| \leq \left\|\frac{1}{qn_1n_2}\mathcal{P}_{T^\perp}\mathcal{A}^\perp\left(\boldsymbol{F}\right)\right\| + \left\|\frac{1}{q}\mathcal{P}_{T^\perp}\left(\boldsymbol{A}_{\alpha_i}\right)\langle\boldsymbol{A}_{\alpha_i},\boldsymbol{F}\rangle\right\| \leq \frac{2\sqrt{\nu(\boldsymbol{F})}}{q}.$$

Applying the Operator Bernstein inequality [28, Theorem 6] yields that for any $t \leq \sqrt{\nu(\boldsymbol{F})}n_1n_2$, we have

$$\left\|\mathcal{P}_{T^\perp}\left(\frac{n_1n_2}{m}\mathcal{A}_\Omega + \mathcal{A}^\perp\right)(\boldsymbol{F})\right\| > t$$

with probability at most $c_8\exp\left(-\frac{c_9qt^2}{\nu(\boldsymbol{F})n_1n_2}\right)$ for some positive constants c_8 and c_9.

G Proof of Lemma 8

Suppose there is a non-zero perturbation $(\boldsymbol{H},\boldsymbol{T})$ such that $(\boldsymbol{X}+\boldsymbol{H},\boldsymbol{S}+\boldsymbol{T})$ is the optimizer of Robust-EMaC. One can easily verify that $\mathcal{P}_{\Omega^\perp}(\boldsymbol{S}+\boldsymbol{T}) = 0$, otherwise we can always set $\boldsymbol{S}+\boldsymbol{T}$ as $\mathcal{P}_\Omega(\boldsymbol{S}+\boldsymbol{T})$ to yield a better estimate. This together with the fact that $\mathcal{P}_{\Omega^\perp}(\boldsymbol{S}) = 0$ implies that $\mathcal{P}_\Omega(\boldsymbol{T}) = \boldsymbol{T}$. Observe that the constraints of Robust-EMaC indicate

$$\mathcal{P}_\Omega(\boldsymbol{X}+\boldsymbol{S}) = \mathcal{P}_\Omega(\boldsymbol{X}+\boldsymbol{H}+\boldsymbol{S}+\boldsymbol{T}) \quad \Rightarrow \quad \mathcal{P}_\Omega(\boldsymbol{H}+\boldsymbol{T}) = 0,$$

which is equivalent to requiring $\mathcal{A}'_\Omega(\boldsymbol{H}_\mathrm{e}) = -\mathcal{A}'_\Omega(\boldsymbol{T}_\mathrm{e}) = -\boldsymbol{T}_\mathrm{e}$ and $\mathcal{A}^\perp(\boldsymbol{H}_\mathrm{e}) = 0$.

Recall that $\boldsymbol{H}_\mathrm{e}$ and $\boldsymbol{S}_\mathrm{e}$ are the enhanced forms of $\boldsymbol{H}$ and $\boldsymbol{S}$, respectively. Set $\boldsymbol{W}_0 \in T^\perp$ to be a matrix satisfying $\langle\boldsymbol{W}_0,\mathcal{P}_{T^\perp}(\boldsymbol{H}_\mathrm{e})\rangle = \|\mathcal{P}_{T^\perp}(\boldsymbol{H}_\mathrm{e})\|_*$ and $\|\boldsymbol{W}_0\| \leq 1$, then $\boldsymbol{U}\boldsymbol{V}^*+\boldsymbol{W}_0$ is a subgradient of the nuclear norm at $\boldsymbol{X}_\mathrm{e}$. This gives

$$\begin{aligned}
\|\boldsymbol{X}_\mathrm{e}+\boldsymbol{H}_\mathrm{e}\|_* &\geq \|\boldsymbol{X}_\mathrm{e}\|_* + \langle\boldsymbol{U}\boldsymbol{V}^*+\boldsymbol{W}_0,\boldsymbol{H}_\mathrm{e}\rangle \\
&= \|\boldsymbol{X}_\mathrm{e}\|_* + \langle\boldsymbol{U}\boldsymbol{V}^*,\boldsymbol{H}_\mathrm{e}\rangle + \|\mathcal{P}_{T^\perp}(\boldsymbol{H}_\mathrm{e})\|_*.
\end{aligned} \tag{76}$$

Owing to the fact that $\text{support}(\boldsymbol{S}) \subseteq \Omega^{\text{dirty}}$, one has $\boldsymbol{S}_{\text{e}} = \mathcal{A}'_{\Omega^{\text{dirty}}}(\boldsymbol{S}_{\text{e}})$. Combining this and the fact that $\text{support}(\boldsymbol{S}_{\text{e}} + \boldsymbol{T}_{\text{e}}) \subseteq \Omega$ yields

$$\|\boldsymbol{S}_{\text{e}} + \boldsymbol{T}_{\text{e}}\|_1 = \|\mathcal{A}'_{\Omega^{\text{clean}}}(\boldsymbol{T}_{\text{e}})\|_1 + \|\boldsymbol{S}_{\text{e}} + \mathcal{A}'_{\Omega^{\text{dirty}}}(\boldsymbol{T}_{\text{e}})\|_1,$$

which further gives

$$\begin{aligned}
\|\boldsymbol{S}_{\text{e}} + \boldsymbol{T}_{\text{e}}\|_1 - \|\boldsymbol{S}_{\text{e}}\|_1 &= \|\mathcal{A}'_{\Omega^{\text{clean}}}(\boldsymbol{T}_{\text{e}})\|_1 + \|\boldsymbol{S}_{\text{e}} + \mathcal{A}'_{\Omega^{\text{dirty}}}(\boldsymbol{T}_{\text{e}})\|_1 - \|\boldsymbol{S}_{\text{e}}\|_1 \\
&\geq \|\mathcal{A}'_{\Omega^{\text{clean}}}(\boldsymbol{T}_{\text{e}})\|_1 + \langle \text{sgn}(\boldsymbol{S}_{\text{e}}), \mathcal{A}'_{\Omega^{\text{dirty}}}(\boldsymbol{T}_{\text{e}}) \rangle & (77) \\
&= \|\mathcal{A}'_{\Omega^{\text{clean}}}(\boldsymbol{T}_{\text{e}})\|_1 - \langle \text{sgn}(\boldsymbol{S}_{\text{e}}), \mathcal{A}'_{\Omega^{\text{dirty}}}(\boldsymbol{H}_{\text{e}}) \rangle & (78) \\
&= \|\mathcal{A}'_{\Omega^{\text{clean}}}(\boldsymbol{T}_{\text{e}})\|_1 - \langle \mathcal{A}'_{\Omega^{\text{dirty}}}(\text{sgn}(\boldsymbol{S}_{\text{e}})), \boldsymbol{H}_{\text{e}} \rangle \\
&= \|\mathcal{A}'_{\Omega^{\text{clean}}}(\boldsymbol{H}_{\text{e}})\|_1 - \langle \text{sgn}(\boldsymbol{S}_{\text{e}}), \boldsymbol{H}_{\text{e}} \rangle. & (79)
\end{aligned}$$

Here, (77) follows from the fact that $\text{sgn}(\boldsymbol{S}_{\text{e}})$ is the subgradient of $\|\cdot\|_1$ at $\boldsymbol{S}_{\text{e}}$, and (78) arises from the identity $\mathcal{P}_{\Omega^{\text{dirty}}}(\boldsymbol{H} + \boldsymbol{T}) = 0$ and hence $\mathcal{A}'_{\Omega^{\text{dirty}}}(\boldsymbol{H}_{\text{e}}) = -\mathcal{A}'_{\Omega^{\text{dirty}}}(\boldsymbol{T}_{\text{e}})$. The inequalities (76) and (79) taken collectively lead to

$$\begin{aligned}
&\|\boldsymbol{X}_{\text{e}} + \boldsymbol{H}_{\text{e}}\|_* + \lambda \|\boldsymbol{S}_{\text{e}} + \boldsymbol{T}_{\text{e}}\|_1 - (\|\boldsymbol{X}_{\text{e}}\|_* + \lambda \|\boldsymbol{S}_{\text{e}}\|_1) \\
\geq\ & \langle \boldsymbol{U}\boldsymbol{V}^*, \boldsymbol{H}_{\text{e}} \rangle + \|\mathcal{P}_{T^\perp}(\boldsymbol{H}_{\text{e}})\|_* + \lambda \|\mathcal{A}'_{\Omega^{\text{clean}}}(\boldsymbol{H}_{\text{e}})\|_1 - \lambda \langle \text{sgn}(\boldsymbol{S}_{\text{e}}), \boldsymbol{H}_{\text{e}} \rangle \\
\geq & -\langle \lambda \text{sgn}(\boldsymbol{S}_{\text{e}}) - \boldsymbol{U}\boldsymbol{V}^*, \boldsymbol{H}_{\text{e}} \rangle + \|\mathcal{P}_{T^\perp}(\boldsymbol{H}_{\text{e}})\|_* + \lambda \|\mathcal{A}'_{\Omega^{\text{clean}}}(\boldsymbol{H}_{\text{e}})\|_1. \qquad (80)
\end{aligned}$$

It remains to show that the right-hand side of (80) cannot be negative. For a dual matrix $\boldsymbol{W}$ satisfying Conditions (44), one can derive

$$\begin{aligned}
&\langle \lambda \text{sgn}(\boldsymbol{S}_{\text{e}}) - \boldsymbol{U}\boldsymbol{V}^*, \boldsymbol{H}_{\text{e}} \rangle \\
=& \langle \boldsymbol{W} + \lambda \text{sgn}(\boldsymbol{S}_{\text{e}}) - \boldsymbol{U}\boldsymbol{V}^*, \boldsymbol{H}_{\text{e}} \rangle - \langle \boldsymbol{W}, \boldsymbol{H}_{\text{e}} \rangle \\
=& \langle \mathcal{P}_T(\boldsymbol{W} + \lambda \text{sgn}(\boldsymbol{S}_{\text{e}}) - \boldsymbol{U}\boldsymbol{V}^*), \mathcal{P}_T(\boldsymbol{H}_{\text{e}}) \rangle + \langle \mathcal{P}_{T^\perp}(\boldsymbol{W} + \lambda \text{sgn}(\boldsymbol{S}_{\text{e}}) - \boldsymbol{U}\boldsymbol{V}^*), \mathcal{P}_{T^\perp}(\boldsymbol{H}_{\text{e}}) \rangle \\
&- \langle \mathcal{A}'_{\Omega^{\text{clean}}}(\boldsymbol{W}), \mathcal{A}'_{\Omega^{\text{clean}}}(\boldsymbol{H}_{\text{e}}) \rangle - \left\langle \mathcal{A}'_{(\Omega^{\text{clean}})^\perp}(\boldsymbol{W}), \mathcal{A}'_{(\Omega^{\text{clean}})^\perp}(\boldsymbol{H}_{\text{e}}) \right\rangle. \\
\leq& \frac{\lambda}{n_1^2 n_2^2} \|\mathcal{P}_T(\boldsymbol{H}_{\text{e}})\|_{\text{F}} + \frac{1}{4} \|\mathcal{P}_{T^\perp}(\boldsymbol{H}_{\text{e}})\|_* + \frac{\lambda}{4} \|\mathcal{A}'_{\Omega^{\text{clean}}}(\boldsymbol{H}_{\text{e}})\|_1, \qquad (81)
\end{aligned}$$

where the last inequality follows from the four properties of $\boldsymbol{W}$ in (44). Since $(\boldsymbol{X} + \boldsymbol{H}, \boldsymbol{S} + \boldsymbol{T})$ is assumed to be the optimizer, substituting (81) into (80) then yields

$$\begin{aligned}
0 &\geq \|\boldsymbol{X}_{\text{e}} + \boldsymbol{H}_{\text{e}}\|_* + \lambda \|\boldsymbol{S}_{\text{e}} + \boldsymbol{T}_{\text{e}}\|_1 - (\|\boldsymbol{X}_{\text{e}}\|_* + \lambda \|\boldsymbol{S}_{\text{e}}\|_1) & (82) \\
&\geq \frac{3}{4} \|\mathcal{P}_{T^\perp}(\boldsymbol{H}_{\text{e}})\|_* + \frac{3}{4}\lambda \|\mathcal{A}'_{\Omega^{\text{clean}}}(\boldsymbol{H}_{\text{e}})\|_1 - \frac{\lambda}{n_1^2 n_2^2} \|\mathcal{P}_T(\boldsymbol{H}_{\text{e}})\|_{\text{F}} \\
&\geq \frac{3}{4} \|\mathcal{P}_{T^\perp}(\boldsymbol{H}_{\text{e}})\|_* + \frac{3}{4}\lambda \|\mathcal{A}'_{\Omega^{\text{clean}}}(\boldsymbol{H}_{\text{e}})\|_{\text{F}} - \frac{\lambda}{n_1^2 n_2^2} \|\mathcal{P}_T(\boldsymbol{H}_{\text{e}})\|_{\text{F}}, & (83)
\end{aligned}$$

where (83) arises due to the inequality $\|\boldsymbol{M}\|_{\text{F}} \leq \|\boldsymbol{M}\|_1$.

The invertibility condition (42) on $\mathcal{P}_T \mathcal{A}_{\Omega^{\text{clean}}} \mathcal{P}_T$ is equivalent to

$$\left\| \mathcal{P}_T - \mathcal{P}_T \left(\frac{1}{\rho(1-s)} \mathcal{A}_{\Omega^{\text{clean}}} + \mathcal{A}^\perp \right) \mathcal{P}_T \right\| \leq \frac{1}{2},$$

indicating that

$$\frac{1}{2} \|\mathcal{P}_T(\boldsymbol{H}_{\text{e}})\|_{\text{F}} \leq \left\| \mathcal{P}_T \left(\frac{1}{\rho(1-s)} \mathcal{A}_{\Omega^{\text{clean}}} + \mathcal{A}^\perp \right) \mathcal{P}_T(\boldsymbol{H}_{\text{e}}) \right\|_{\text{F}} \leq \frac{3}{2} \|\mathcal{P}_T(\boldsymbol{H}_{\text{e}})\|_{\text{F}}.$$

One can, therefore, bound $\left\|\mathcal{P}_{T}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\|_{\mathrm{F}}$ as follows

$$
\begin{aligned}
\left\|\mathcal{P}_{T}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\|_{\mathrm{F}} &\leq 2\left\|\mathcal{P}_{T}\left(\frac{1}{\rho(1-s)}\mathcal{A}_{\Omega^{\text{clean}}}+\mathcal{A}^{\perp}\right)\mathcal{P}_{T}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\|_{\mathrm{F}} \\
&\leq \frac{2}{\rho(1-s)}\left\|\mathcal{P}_{T}\mathcal{A}_{\Omega^{\text{clean}}}\mathcal{P}_{T}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\|_{\mathrm{F}}+2\left\|\mathcal{P}_{T}\mathcal{A}^{\perp}\mathcal{P}_{T}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\|_{\mathrm{F}} \\
&\leq \frac{2}{\rho(1-s)}\left(\left\|\mathcal{P}_{T}\mathcal{A}_{\Omega^{\text{clean}}}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\|_{\mathrm{F}}+\left\|\mathcal{P}_{T}\mathcal{A}_{\Omega^{\text{clean}}}\mathcal{P}_{T^{\perp}}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\|_{\mathrm{F}}\right) \\
&\quad +2\left\|\mathcal{P}_{T}\mathcal{A}^{\perp}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\|_{\mathrm{F}}+2\left\|\mathcal{P}_{T}\mathcal{A}^{\perp}\mathcal{P}_{T^{\perp}}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\|_{\mathrm{F}} \\
&\leq \frac{2}{\rho(1-s)}\left(\left\|\mathcal{A}_{\Omega^{\text{clean}}}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\|_{\mathrm{F}}+\left\|\mathcal{A}_{\Omega^{\text{clean}}}\mathcal{P}_{T^{\perp}}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\|_{\mathrm{F}}\right)+2\left\|\mathcal{P}_{T^{\perp}}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\|_{\mathrm{F}}, \qquad (84)
\end{aligned}
$$

where the last inequality exploit the facts that $\mathcal{A}^{\perp}\left(\boldsymbol{H}_{\mathrm{e}}\right)=0$ and $\left\|\mathcal{P}_{T}(\boldsymbol{M})\right\|_{\mathrm{F}}\leq\|\boldsymbol{M}\|_{\mathrm{F}}$.

Recall that $\mathcal{A}_{\Omega^{\text{clean}}}$ corresponds to sampling with replacement. Condition (43) together with (84) leads to

$$
\begin{aligned}
\left\|\mathcal{P}_{T}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\|_{\mathrm{F}} &\leq \frac{20\log\left(n_{1}n_{2}\right)}{\rho(1-s)}\left(\left\|\mathcal{A}_{\Omega^{\text{clean}}}^{\prime}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\|_{\mathrm{F}}+\left\|\mathcal{A}_{\Omega^{\text{clean}}}^{\prime}\mathcal{P}_{T^{\perp}}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\|_{\mathrm{F}}\right)+2\left\|\mathcal{P}_{T^{\perp}}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\|_{\mathrm{F}} \\
&\leq \frac{20\log\left(n_{1}n_{2}\right)}{\rho(1-s)}\left\|\mathcal{A}_{\Omega^{\text{clean}}}^{\prime}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\|_{\mathrm{F}}+\left(\frac{20\log\left(n_{1}n_{2}\right)}{\rho(1-s)}+2\right)\left\|\mathcal{P}_{T^{\perp}}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\|_{\mathrm{F}} \\
&\leq \frac{20\log\left(n_{1}n_{2}\right)}{\rho(1-s)}\left\|\mathcal{A}_{\Omega^{\text{clean}}}^{\prime}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\|_{\mathrm{F}}+\left(\frac{20\log\left(n_{1}n_{2}\right)}{\rho(1-s)}+2\right)\left\|\mathcal{P}_{T^{\perp}}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\|_{*}, \qquad (85)
\end{aligned}
$$

where the last inequality follows from the fact that $\|\boldsymbol{M}\|_{\mathrm{F}}\leq\|\boldsymbol{M}\|_{*}$. Substituting (85) into (83) yields

$$
\left(\frac{3}{4}-\frac{\lambda}{n_{1}^{2}n_{2}^{2}}\left(\frac{20\log\left(n_{1}n_{2}\right)}{\rho(1-s)}+2\right)\right)\left\|\mathcal{P}_{T^{\perp}}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\|_{*}+\lambda\left(\frac{3}{4}-\frac{20\log\left(n_{1}n_{2}\right)}{\rho(1-s)n_{1}^{2}n_{2}^{2}}\right)\left\|\mathcal{A}_{\Omega^{\text{clean}}}^{\prime}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\|_{\mathrm{F}}\leq 0. \qquad (86)
$$

Since $\lambda<1$ and $\rho n_{1}^{2}n_{2}^{2}\gg\log\left(n_{1}n_{2}\right)$, both terms on the left-hand side of (86) are positive. This can only occur when

$$
\mathcal{P}_{T^{\perp}}\left(\boldsymbol{H}_{\mathrm{e}}\right)=0 \quad \text{and} \quad \mathcal{A}_{\Omega^{\text{clean}}}^{\prime}\left(\boldsymbol{H}_{\mathrm{e}}\right)=0. \qquad (87)
$$

(1) Consider first the situation where

$$
\left\|\mathcal{P}_{T}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\|_{\mathrm{F}}\leq\frac{n_{1}^{2}n_{2}^{2}}{2}\left\|\mathcal{P}_{T^{\perp}}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\|_{\mathrm{F}}. \qquad (88)
$$

One can immediately see that

$$
\begin{aligned}
&\left\|\mathcal{P}_{T}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\|_{\mathrm{F}}\leq\frac{n_{1}^{2}n_{2}^{2}}{2}\left\|\mathcal{P}_{T^{\perp}}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\|_{\mathrm{F}}=0 \\
\Rightarrow\quad &\mathcal{P}_{T}\left(\boldsymbol{H}_{\mathrm{e}}\right)=\mathcal{P}_{T^{\perp}}\left(\boldsymbol{H}_{\mathrm{e}}\right)=0;\quad\Rightarrow\quad \boldsymbol{H}_{\mathrm{e}}=0.
\end{aligned}
$$

That said, Robust-EMaC succeeds in finding $\boldsymbol{X}_{\mathrm{e}}$ under Condition (88).

(2) Consider instead the complement situation where

$$
\left\|\mathcal{P}_{T}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\|_{\mathrm{F}}>\frac{n_{1}^{2}n_{2}^{2}}{2}\left\|\mathcal{P}_{T^{\perp}}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\|_{\mathrm{F}}.
$$

Note that $\mathcal{A}_{\Omega^{\text{clean}}}^{\prime}(\boldsymbol{H}_{\mathrm{e}})=\mathcal{A}^{\perp}(\boldsymbol{H}_{\mathrm{e}})=0$ and $\left\|\mathcal{P}_{T}\mathcal{A}\mathcal{P}_{T}-\frac{1}{\rho(1-s)}\mathcal{P}_{T}\mathcal{A}_{\Omega^{\text{clean}}}\mathcal{P}_{T}\right\|\leq\frac{1}{2}$. Using the same argument as in the proof of Lemma 3 (see the second part of Appendix B) with Ω replaced by Ω^{clean}, we can conclude $\boldsymbol{H}_{\mathrm{e}}=0$.

H Proof of Lemma 9

By definition of $\nu(\cdot)$, we can bound

$$\begin{aligned}
\nu\left(\boldsymbol{F}_0\right) &= \max_{(k,l)\in[n_1]\times[n_2]} \frac{1}{\omega_{k,l}} \left|\left\langle \boldsymbol{A}_{(k,l)}, \boldsymbol{U}\boldsymbol{V}^* - \lambda \mathcal{P}_T \mathrm{sgn}\left(\boldsymbol{S}_{\mathrm{e}}\right)\right\rangle\right|^2 \\
&\leq \max_{(k,l)\in[n_1]\times[n_2]} \left\{ \frac{2}{\omega_{k,l}} \left|\left\langle \boldsymbol{A}_{(k,l)}, \boldsymbol{U}\boldsymbol{V}^*\right\rangle\right|^2 + \frac{2}{\omega_{k,l}} \lambda^2 \left|\left\langle \boldsymbol{A}_{(k,l)}, \mathcal{P}_T\left(\mathrm{sgn}\left(\boldsymbol{S}_{\mathrm{e}}\right)\right)\right\rangle\right|^2 \right\} \\
&\leq \frac{2\mu_2 r}{n_1^2 n_2^2} + \max_{(k,l)\in[n_1]\times[n_2]} \frac{2\lambda^2}{\omega_{k,l}} \left|\left\langle \boldsymbol{A}_{(k,l)}, \mathcal{P}_T\left(\mathrm{sgn}\left(\boldsymbol{S}_{\mathrm{e}}\right)\right)\right\rangle\right|^2, \qquad (89)
\end{aligned}$$

where the last inequality follows from the definition of μ_2 in (19). In order to develop a good bound on $\nu\left(\boldsymbol{F}_0\right)$, we still need to bound the component $\left|\left\langle \boldsymbol{A}_{(k,l)}, \mathcal{P}_T\left(\mathrm{sgn}\left(\boldsymbol{S}_{\mathrm{e}}\right)\right)\right\rangle\right|$. This is achieved through the following lemma.

Lemma 11. *Suppose that s is a positive constant. If $\rho > c_{21}\frac{\max(\mu_4,\mu_1 c_{\mathrm{s}})r}{n_1 n_2}\log\left(n_1 n_2\right)$, one has*

$$\max_{(k,l)\in[n_1]\times[n_2]} \frac{1}{\omega_{k,l}} \left|\left\langle \boldsymbol{A}_{(k,l)}, \mathcal{P}_T\left(\mathrm{sgn}\left(\boldsymbol{S}_{\mathrm{e}}\right)\right)\right\rangle\right|^2 \leq c_{20} \frac{\rho s \max\left(\mu_4, \mu_1 c_{\mathrm{s}}\right) r \log\left(n_1 n_2\right)}{n_1 n_2}$$

with probability at least $1-\left(n_1 n_2\right)^{-4}$.

By setting $\lambda := \frac{1}{\sqrt{n_1 n_2 \rho \log(n_1 n_2)}}$, we can derive from Lemma 11 that

$$\max_{(k,l)\in[n_1]\times[n_2]} \frac{2}{\omega_{k,l}} \lambda^2 \left|\left\langle \boldsymbol{A}_{(k,l)}, \mathcal{P}_T\left(\mathrm{sgn}\left(\boldsymbol{S}_{\mathrm{e}}\right)\right)\right\rangle\right|^2 < c_{20} \frac{\max\left(\mu_4, \mu_1 c_{\mathrm{s}}\right) r}{n_1^2 n_2^2},$$

where we also use the assumption $s \leq 1/2$. This together with (89) leads to

$$\nu\left(\boldsymbol{F}_0\right) \leq \frac{c_9 \max\left(\mu_1 c_{\mathrm{s}}, \mu_2, \mu_4\right) r}{n_1^2 n_2^2}$$

for some $c_9 > 0$. Applying Lemma 6 then shows that

$$\nu\left(\boldsymbol{F}_i\right) \leq \frac{1}{4^i} \nu\left(\boldsymbol{F}_0\right) \leq \frac{c_9 \max\left(\mu_1 c_{\mathrm{s}}, \mu_4\right) r}{4^i n_1^2 n_2^2} \qquad (90)$$

with high probability, as long as $m > c_7 \max\left(\mu_1 c_{\mathrm{s}}, \mu_4\right) r \log^2\left(n_1 n_2\right)$ for some $c_7 > 0$.

Finally, the proof of Lemma 11 proceeds as follows.

Proof. By definition, Ω^{dirty} is the set of *distinct* locations that appear in Ω but not in Ω^{clean}. To simplify the analysis, we introduce an auxiliary multiset $\tilde{\Omega}^{\mathrm{dirty}}$ that contains $\rho s n_1 n_2$ i.i.d. entries. Specifically, suppose that $\Omega = \{\boldsymbol{a}_i \mid 1 \leq i \leq \rho n_1 n_2\}$, $\Omega^{\mathrm{clean}} = \{\boldsymbol{a}_i \mid 1 \leq i \leq \rho\left(1-s\right) n_1 n_2\}$ and $\tilde{\Omega}^{\mathrm{dirty}} = \{\boldsymbol{a}_i \mid \rho\left(1-s\right) n_1 n_2 < i \leq \rho n_1 n_2\}$, where $\boldsymbol{a}_i$'s are independently and uniformly selected from $[n_1]\times[n_2]$.

In addition, we consider an equivalent model for $\mathrm{sgn}\left(\boldsymbol{S}\right)$ as follows

- Define $\boldsymbol{K} = \left(K_{\alpha,\beta}\right)_{1\leq\alpha\leq n_1, 1\leq\beta\leq n_2}$ to be a random $n_1 \times n_2$ matrix such that all of its entries are independent and have amplitude 1 (i.e. in the real case, all entries are either 1 or -1, and in the complex case, all entries have amplitude 1 and arbitrary phase). We assume that $\mathbb{E}\boldsymbol{K} = \boldsymbol{0}$.
- Set $\mathrm{sgn}\left(\boldsymbol{S}\right)$ such that $\mathrm{sgn}\left(\boldsymbol{S}_{\alpha,\beta}\right) = K_{\alpha,\beta}\boldsymbol{1}_{\{(\alpha,\beta)\in\Omega^{\mathrm{dirty}}\}}$, and hence

$$\mathrm{sgn}\left(\boldsymbol{S}_{\mathrm{e}}\right) = \sum_{(\alpha,\beta)\in\Omega^{\mathrm{dirty}}} K_{\alpha,\beta}\sqrt{\omega_{\alpha,\beta}}\boldsymbol{A}_{\alpha,\beta}.$$

Recall that $\operatorname{support}(\boldsymbol{S}) \subseteq \Omega^{\text{dirty}}$. Rather than directly studying the property of $\operatorname{sgn}(\boldsymbol{S}_{\text{e}})$, we will first examine the property of an auxiliary matrix

$$\tilde{\boldsymbol{S}}_{\text{e}} := \sum_{i=\rho(1-s)n_1n_2+1}^{\rho n_1 n_2} K_{\boldsymbol{a}_i}\sqrt{\omega_{\boldsymbol{a}_i}}\boldsymbol{A}_{\boldsymbol{a}_i},$$

and then bound the difference between $\tilde{\boldsymbol{S}}_{\text{e}}$ and $\operatorname{sgn}(\boldsymbol{S}_{\text{e}})$.

For any given pair $(k,l) \in [n_1] \times [n_2]$, define a random variable

$$\mathcal{Z}_{\alpha,\beta} := \sqrt{\frac{\omega_{\alpha,\beta}}{\omega_{k,l}}}\left\langle \mathcal{P}_T \boldsymbol{A}_{(k,l)}, K_{\alpha,\beta}\boldsymbol{A}_{\alpha,\beta}\right\rangle.$$

Thus, $\mathcal{Z}_{\boldsymbol{a}_i}$'s are conditionally independent given $\boldsymbol{K}$. The conditional mean and second moment of $\mathcal{Z}_{\boldsymbol{a}_i}$ can be computed as

$$\mathbb{E}\left(\mathcal{Z}_{\boldsymbol{a}_i} \mid \boldsymbol{K}\right) = \frac{1}{n_1n_2}\frac{1}{\sqrt{\omega_{k,l}}}\left\langle \mathcal{P}_T \boldsymbol{A}_{(k,l)}, \boldsymbol{K}_{\text{e}}\right\rangle,$$

and

$$\begin{aligned}\mathbf{Var}\left(\mathcal{Z}_{\boldsymbol{a}_i} \mid \boldsymbol{K}\right) \leq \mathbb{E}\left(\mathcal{Z}_{\boldsymbol{a}_i}\mathcal{Z}_{\boldsymbol{a}_i}^* \mid \boldsymbol{K}\right) &= \frac{1}{n_1n_2}\frac{1}{\omega_{k,l}}\sum_{\boldsymbol{b}\in[n_1]\times[n_2]}\omega_{\boldsymbol{b}}\left|\left\langle \mathcal{P}_T \boldsymbol{A}_{(k,l)}, \boldsymbol{A}_{\boldsymbol{b}}\right\rangle\right|^2 \\ &\leq \frac{\max\left(\mu_1 c_{\text{s}}, \mu_4\right) r}{n_1^2 n_2^2} := V,\end{aligned}$$

where the last inequality follows from the definition of μ_4. Besides, applying Lemma 2 allows us to bound the magnitude of $\mathcal{Z}_{\alpha,\beta}$ as follows

$$\left|\mathcal{Z}_{\alpha,\beta}\right| \leq \frac{3\mu_1 c_{\text{s}} r}{n_1 n_2} := B. \tag{91}$$

Applying the Bernstein inequality [28, Theorem 6] yields that: for any $t \leq \frac{2}{3}\rho s$,

$$\begin{aligned}\mathbb{P}\left(\left|\left(\sum_{i=\rho(1-s)n_1n_2+1}^{\rho n_1n_2}\mathcal{Z}_{\boldsymbol{a}_i}\right) - \frac{\rho s}{\sqrt{\omega_{k,l}}}\left\langle \mathcal{P}_T \boldsymbol{A}_{(k,l)}, \boldsymbol{K}_{\text{e}}\right\rangle\right| > t\right) &\leq 2n_1n_2\exp\left(-\frac{t^2}{4\rho s n_1 n_2 V}\right) \\ &= 2n_1n_2\exp\left(-\frac{t^2}{\frac{4\rho s\max(\mu_1 c_{\text{s}},\mu_4) r}{n_1n_2}}\right).\end{aligned} \tag{92}$$

Recall that $\tilde{\boldsymbol{S}}_{\text{e}} := \sum_{i=\rho(1-s)n_1n_2+1}^{\rho n_1n_2} K_{\boldsymbol{a}_i}\sqrt{\omega_{\boldsymbol{a}_i}}\boldsymbol{A}_{\boldsymbol{a}_i}$. Conditional on $\boldsymbol{K}$, $\frac{1}{\sqrt{\omega_{k,l}}}\left\langle \boldsymbol{A}_{(k,l)}, \mathcal{P}_T\left(\tilde{\boldsymbol{S}}_{\text{e}}\right)\right\rangle$ is equivalent to $\sum_{i=\rho(1-s)n_1n_2+1}^{\rho n_1n_2}\mathcal{Z}_{\boldsymbol{a}_i}$ in distribution. We also note that for any constant $c_{20} > 0$, there exists a constant $c_{21} > 0$ such that if $\rho s n_1 n_2 > c_{21}\max\left(\mu_1 c_{\text{s}}, \mu_4\right) r \log\left(n_1n_2\right)$, then one has

$$\sqrt{\frac{c_{20}\rho s\max\left(\mu_1 c_{\text{s}}, \mu_4\right) r\log\left(n_1 n_2\right)}{n_1n_2}} \leq \frac{2}{3}\rho s.$$

Therefore, by setting $t := \sqrt{\frac{c_{20}\rho s\max(\mu_1 c_{\text{s}},\mu_4) r\log(n_1n_2)}{n_1n_2}}$ in (92), we can show that with probability exceeding $1 - \left(n_1n_2\right)^{-4}$, one has, when conditional on $\boldsymbol{K}$, that

$$\frac{1}{\omega_{k,l}}\left|\left\langle \boldsymbol{A}_{(k,l)}, \mathcal{P}_T\left(\tilde{\boldsymbol{S}}_{\text{e}}\right)\right\rangle - \rho s\left\langle \mathcal{P}_T \boldsymbol{A}_{(k,l)}, \boldsymbol{K}_{\text{e}}\right\rangle\right|^2 \leq \frac{c_{20}\rho s\max\left(\mu_1 c_{\text{s}}, \mu_4\right) r\log\left(n_1n_2\right)}{n_1n_2} \tag{93}$$

for every $(k,l) \in [n_1] \times [n_2]$, provided that $\rho s n_1 n_2 > c_{21}\max\left(\mu_1 c_{\text{s}}, \mu_4\right) r\log\left(n_1n_2\right)$ for some constants $c_{20}, c_{21} > 0$.

The next step is to bound $\frac{\rho s}{\sqrt{\omega_{k,l}}}\left\langle \mathcal{P}_T \boldsymbol{A}_{(k,l)}, \boldsymbol{K}_{\mathrm{e}}\right\rangle$. For convenience of analysis, we represent as

$$\boldsymbol{K}_{\mathrm{e}} = \sum_{\boldsymbol{a}\in[n_1]\times[n_2]} z_{\boldsymbol{a}}\sqrt{\omega_{\boldsymbol{a}}}\boldsymbol{A}_{\boldsymbol{a}},$$

where $z_{\boldsymbol{a}}$'s are independent (not necessarily i.i.d.) zero-mean random variables satisfying $|z_{\boldsymbol{a}}| = 1$. Let $\mathcal{Y}_{\boldsymbol{a}} := \frac{1}{\sqrt{\omega_{k,l}}}\left\langle \mathcal{P}_T \boldsymbol{A}_{(k,l)}, z_{\boldsymbol{a}}\sqrt{\omega_{\boldsymbol{a}}}\boldsymbol{A}_{\boldsymbol{a}}\right\rangle$, then the definition of μ_4 and Lemma 2 allow us to compute

$$\mathbb{E}\mathcal{Y}_{\boldsymbol{a}} = 0,$$

$$|\mathcal{Y}_{\boldsymbol{a}}| = \frac{1}{\sqrt{\omega_{k,l}}}\left|\left\langle \mathcal{P}_T \boldsymbol{A}_{(k,l)}, \sqrt{\omega_{\boldsymbol{a}}}\boldsymbol{A}_{\boldsymbol{a}}\right\rangle\right| \leq \frac{3\mu_1 c_{\mathrm{s}} r}{n_1 n_2},$$

and

$$\sum_{\boldsymbol{a}\in[n_1]\times[n_2]} \mathbb{E}\mathcal{Y}_{\boldsymbol{a}}\mathcal{Y}_{\boldsymbol{a}}^* = \frac{1}{\omega_{k,l}}\sum_{\boldsymbol{a}\in[n_1]\times[n_2]}\left|\left\langle \mathcal{P}_T \boldsymbol{A}_{(k,l)}, \sqrt{\omega_{\boldsymbol{a}}}\boldsymbol{A}_{\boldsymbol{a}}\right\rangle\right|^2 \leq \frac{\max\left\{\mu_1 c_{\mathrm{s}}, \mu_4\right\} r}{n_1 n_2} := \tilde{V}.$$

Applying the Bernstein inequality [28, Theorem 6] suggests that for any $t < \frac{2}{3}$,

$$\mathbb{P}\left\{\left|\frac{1}{\sqrt{\omega_{k,l}}}\left\langle \mathcal{P}_T \boldsymbol{A}_{(k,l)}, \boldsymbol{K}_{\mathrm{e}}\right\rangle\right| \geq t\right\} \leq 2n_1 n_2 \exp\left(-\frac{t^2}{4\frac{\max\{\mu_1 c_{\mathrm{s}},\mu_4\} r}{n_1 n_2}}\right).$$

Consequently, there exists a constant $c_{24} > 0$ such that

$$\frac{\rho^2 s^2}{\omega_{k,l}}\left|\left\langle \mathcal{P}_T \boldsymbol{A}_{(k,l)}, \boldsymbol{K}_{\mathrm{e}}\right\rangle\right|^2 \leq \frac{c_{24}\rho^2 s^2 \max\left\{\mu_1 c_{\mathrm{s}}, \mu_4\right\} r \log\left(n_1 n_2\right)}{n_1 n_2}$$

with high probability. This together with (93) suggests that

$$\begin{aligned}
\frac{1}{\omega_{k,l}}\left|\left\langle \boldsymbol{A}_{(k,l)}, \mathcal{P}_T\left(\tilde{\boldsymbol{S}}_{\mathrm{e}}\right)\right\rangle\right|^2 &= \left|\sum_{i=\rho(1-s)n_1 n_2+1}^{\rho n_1 n_2} \mathcal{Z}_{\boldsymbol{a}_i}\right|^2 \\
&\leq 2\left|\left(\sum_{i=\rho(1-s)n_1 n_2+1}^{\rho n_1 n_2} \mathcal{Z}_{\boldsymbol{a}_i}\right) - \frac{\rho s}{\sqrt{\omega_{k,l}}}\left\langle \mathcal{P}_T \boldsymbol{A}_{(k,l)}, \boldsymbol{K}_{\mathrm{e}}\right\rangle\right|^2 + 2\frac{\rho^2 s^2}{\omega_{k,l}}\left|\left\langle \mathcal{P}_T \boldsymbol{A}_{(k,l)}, \boldsymbol{K}_{\mathrm{e}}\right\rangle\right|^2 \\
&\leq \frac{4c_{24}\rho s \max\left(\mu_1 c_{\mathrm{s}}, \mu_4\right) r \log\left(n_1 n_2\right)}{n_1 n_2} \qquad (94)
\end{aligned}$$

with high probability.

We still need to bound the deviation of $\tilde{\boldsymbol{S}}_{\mathrm{e}}$ from $\mathrm{sgn}\left(\boldsymbol{S}_{\mathrm{e}}\right)$. Observe that the difference between them arise from sampling with replacement, i.e. there are a few entries in $\{\boldsymbol{a}_i \mid \rho\left(1-s\right) n_1 n_2 < i \leq \rho n_1 n_2\}$ that either fall within Ω^{clean} or have appeared more than once. A simple Chernoff bound argument (e.g. [43]) indicates the number of aforementioned conflicts is upper bounded by $10\log\left(n_1 n_2\right)$ with high probability. That said, one can find a collection of entry locations $\{\boldsymbol{b}_1, \cdots, \boldsymbol{b}_N\}$ such that

$$\tilde{\boldsymbol{S}}_{\mathrm{e}} - \mathrm{sgn}\left(\boldsymbol{S}_{\mathrm{e}}\right) = \sum_{i=1}^{N} K_{\boldsymbol{b}_i}\sqrt{\omega_{\boldsymbol{b}_i}}\boldsymbol{A}_{\boldsymbol{b}_i}, \qquad (95)$$

where $N \leq 10\log\left(n_1 n_2\right)$ with high probability. Therefore, we can bound

$$\begin{aligned}
\frac{1}{\sqrt{\omega_{k,l}}}\left|\left\langle \boldsymbol{A}_{(k,l)}, \mathcal{P}_T\left(\tilde{\boldsymbol{S}}_{\mathrm{e}} - \mathrm{sgn}\left(\boldsymbol{S}_{\mathrm{e}}\right)\right)\right\rangle\right| &\leq \sum_{i=1}^{N}\frac{1}{\sqrt{\omega_{k,l}}}\left|\left\langle \boldsymbol{A}_{(k,l)}, \mathcal{P}_T\left(\sqrt{\omega_{\boldsymbol{b}_i}}\boldsymbol{A}_{\boldsymbol{b}_i}\right)\right\rangle\right| \\
&\leq N\frac{3\mu_1 c_{\mathrm{s}} r}{n_1 n_2},
\end{aligned}$$

and hence, for some constant $c_{22} > 0$,

$$\frac{1}{\omega_{k,l}} \left| \left\langle \boldsymbol{A}_{(k,l)}, \mathcal{P}_T \left(\tilde{\boldsymbol{S}}_{\text{e}} - \operatorname{sgn}(\boldsymbol{S}_{\text{e}}) \right) \right\rangle \right|^2 \leq c_{22} \frac{\mu_1^2 c_{\text{s}}^2 r^2 \log^2(n_1 n_2)}{n_1^2 n_2^2} \leq \frac{c_{20} \rho s \max(\mu_1 c_{\text{s}}, \mu_4) r \log(n_1 n_2)}{n_1 n_2}$$

holds with high probability, provided that $\rho s n_1 n_2 \geq c_{23} \mu_1 c_{\text{s}} r \log(n_1 n_2)$ for some $c_{23} > 0$.

Putting the above results together yields that for every $(k,l) \in [n_1] \times [n_2]$,

$$\begin{aligned} \frac{1}{\omega_{k,l}} \left| \left\langle \boldsymbol{A}_{(k,l)}, \mathcal{P}_T (\operatorname{sgn}(\boldsymbol{S}_{\text{e}})) \right\rangle \right|^2 &\leq \frac{2}{\omega_{k,l}} \left| \left\langle \boldsymbol{A}_{(k,l)}, \mathcal{P}_T \left(\tilde{\boldsymbol{S}}_{\text{e}} - \operatorname{sgn}(\boldsymbol{S}_{\text{e}}) \right) \right\rangle \right|^2 + \frac{2}{\omega_{k,l}} \left| \left\langle \boldsymbol{A}_{(k,l)}, \mathcal{P}_T \left(\tilde{\boldsymbol{S}}_{\text{e}} \right) \right\rangle \right|^2 \\ &\leq \frac{c_{25} \rho s \max(\mu_1 c_{\text{s}}, \mu_4) r \log(n_1 n_2)}{n_1 n_2} \end{aligned}$$

for some constant $c_{25} > 0$, which completes the proof. □

I Proof of Lemma 10

Consider the model of $\operatorname{sgn}(\boldsymbol{S})$, $\boldsymbol{K}$ and $\tilde{\boldsymbol{S}}_{\text{e}}$ as introduced in the proof of Lemma 11 in Appendix H. For any $(\alpha, \beta) \in [n_1] \times [n_2]$, define

$$\tilde{\mathcal{Z}}_{\alpha,\beta} := \mathcal{A}_{\alpha,\beta}(\boldsymbol{K}_{\text{e}}) = \sqrt{\omega_{\alpha,\beta}} K_{\alpha,\beta} \boldsymbol{A}_{\alpha,\beta}.$$

With this notation, we can see that $\tilde{\mathcal{Z}}_{\boldsymbol{a}_i}$'s are conditionally independent given $\boldsymbol{K}$, and satisfy

$$\mathbb{E}\left(\tilde{\mathcal{Z}}_{\boldsymbol{a}_i} \mid \boldsymbol{K} \right) = \frac{1}{n_1 n_2} \sum_{(\alpha,\beta) \in [n_1] \times [n_2]} \sqrt{\omega_{\alpha,\beta}} \boldsymbol{A}_{\alpha,\beta} K_{\alpha,\beta} = \frac{1}{n_1 n_2} \boldsymbol{K}_{\text{e}},$$

$$\left\| \tilde{\mathcal{Z}}_{\alpha,\beta} \right\| = \left\| \sqrt{\omega_{\alpha,\beta}} \boldsymbol{A}_{\alpha,\beta} \right\| = 1,$$

and

$$\left\| \mathbb{E}\left(\tilde{\mathcal{Z}}_{\boldsymbol{a}_i} \tilde{\mathcal{Z}}_{\boldsymbol{a}_i}^* \mid \boldsymbol{K} \right) \right\| \leq \frac{1}{n_1 n_2} \sum_{(\alpha,\beta) \in [n_1] \times [n_2]} \left\| \omega_{\alpha,\beta} \boldsymbol{A}_{\alpha,\beta} \boldsymbol{A}_{\alpha,\beta}^* \right\| = 1 := V.$$

Since $\tilde{\boldsymbol{S}}_{\text{e}} = \sum_{i=(1-s)\rho n_1 n_2 + 1}^{\rho n_1 n_2} \tilde{\mathcal{Z}}_{\boldsymbol{a}_i}$, applying the Bernstein inequality [28, Theorem 6] implies that for any $t < 2\rho s n_1 n_2$, one has, conditioned on $\boldsymbol{K}$, that

$$\mathbb{P}\left(\left\| \tilde{\boldsymbol{S}}_{\text{e}} - \rho s \boldsymbol{K}_{\text{e}} \right\| > t \right) \leq 2 n_1 n_2 \exp\left(-\frac{t^2}{4 \rho s n_1 n_2 V} \right).$$

Therefore, conditioned on $\boldsymbol{K}$, there exists a constant $c_{22} > 0$ such that

$$\left\| \tilde{\boldsymbol{S}}_{\text{e}} - \rho s \boldsymbol{K}_{\text{e}} \right\| < \sqrt{c_{22} \rho s n_1 n_2 \log(n_1 n_2)} \tag{96}$$

with probability at least than $1 - n_1^{-5} n_2^{-5}$.

The next step is to bound the operator norm of $\rho s \boldsymbol{K}_{\text{e}}$. For convenience of analysis, we write

$$\boldsymbol{K}_{\text{e}} = \sum_{\boldsymbol{a} \in [n_1] \times [n_2]} z_{\boldsymbol{a}} \sqrt{\omega_{\boldsymbol{a}}} \boldsymbol{A}_{\boldsymbol{a}},$$

for a collection of independent (not necessarily i.i.d.) random variables $z_{\boldsymbol{a}}$'s satisfying $|z_{\boldsymbol{a}}| = 1$ and $\mathbb{E} z_{\boldsymbol{a}} = 0$. Let $\mathcal{Y}_{\boldsymbol{a}} := z_{\boldsymbol{a}} \sqrt{\omega_{\boldsymbol{a}}} \boldsymbol{A}_{\boldsymbol{a}}$, then we have $\mathbb{E}\mathcal{Y}_{\boldsymbol{a}} = 0$, $\|\mathcal{Y}_{\boldsymbol{a}}\| = 1$, and

$$\left\| \sum_{\boldsymbol{a} \in [n_1] \times [n_2]} \mathbb{E} \mathcal{Y}_{\boldsymbol{a}} \mathcal{Y}_{\boldsymbol{a}}^* \right\| = \left\| \sum_{\boldsymbol{a} \in [n_1] \times [n_2]} \omega_{\boldsymbol{a}} \boldsymbol{A}_{\boldsymbol{a}} \boldsymbol{A}_{\boldsymbol{a}}^* \right\| \leq n_1 n_2 := \tilde{V}.$$

Therefore, applying the Bernstein inequality yields that for any $t \le n_1 n_2$,

$$\mathbb{P}\left\{\|\boldsymbol{K}_{\mathrm{e}}\| > t\right\} = \mathbb{P}\left\{\left\|\sum_{\boldsymbol{a}\in[n_1]\times[n_2]} \mathcal{Y}_{\boldsymbol{a}}\right\| > t\right\} \le 2n_1 n_2 \exp\left(-\frac{t^2}{4n_1 n_2}\right)$$

or, more simply, there exists a constant $c_{22} > 0$ such that

$$\|\boldsymbol{K}_{\mathrm{e}}\| \le \sqrt{c_{22} n_1 n_2 \log(n_1 n_2)}$$

with high probability. This and (96), taken collectively, yield

$$\left\|\tilde{\boldsymbol{S}}_{\mathrm{e}}\right\| \le \left\|\tilde{\boldsymbol{S}}_{\mathrm{e}} - \rho s \boldsymbol{K}_{\mathrm{e}}\right\| + \rho s \|\boldsymbol{K}_{\mathrm{e}}\| < 2\sqrt{c_{22}\rho s n_1 n_2 \log(n_1 n_2)}$$

with high probability. On the other hand, (95) implies that for sufficiently large n_1 and n_2,

$$\left\|\tilde{\boldsymbol{S}}_{\mathrm{e}} - \mathrm{sgn}(\boldsymbol{S}_{\mathrm{e}})\right\| \le \sum_{i=1}^{N} \left\|\sqrt{\omega_{\boldsymbol{b}_i}} \boldsymbol{A}_{\boldsymbol{b}_i}\right\| = N \le 10\log(n_1 n_2) \ll \sqrt{c_{22}\rho s n_1 n_2 \log(n_1 n_2)}$$

with high probability. Consequently, for a sufficiently small constant s,

$$\begin{aligned}
\|\mathcal{P}_{T^\perp}(\lambda \mathrm{sgn}(\boldsymbol{S}_{\mathrm{e}}))\| \le \lambda \|\mathrm{sgn}(\boldsymbol{S}_{\mathrm{e}})\| &\le \lambda \left\|\tilde{\boldsymbol{S}}_{\mathrm{e}} - \mathrm{sgn}(\boldsymbol{S}_{\mathrm{e}})\right\| + \lambda \left\|\tilde{\boldsymbol{S}}_{\mathrm{e}}\right\| \\
&\le 2\lambda\sqrt{c_{22}\rho s n_1 n_2 \log(n_1 n_2)} \\
&= 2\sqrt{c_{22} s} \le \frac{1}{8}
\end{aligned}$$

with probability exceeding $1 - n_1^{-5} n_2^{-5}$.

J Proof of Theorem 2

We prove this theorem under the conditions of Lemma 3, i.e. (34)–(37). Note that these conditions are satisfied with high probability, as we have shown in the proof of Theorem 1.

Denote the solution of Noisy-EMaC as $\hat{\boldsymbol{X}}_{\mathrm{e}} = \boldsymbol{X}_{\mathrm{e}} + \boldsymbol{H}_{\mathrm{e}}$. Since $\boldsymbol{H}_{\mathrm{e}}$ is a two-fold Hankel matrix, i.e. $\boldsymbol{H}_{\mathrm{e}} = \mathcal{A}_{\Omega}(\boldsymbol{H}_{\mathrm{e}}) + \mathcal{A}_{\Omega^\perp}(\boldsymbol{H}_{\mathrm{e}})$, we can obtain

$$\|\boldsymbol{X}_{\mathrm{e}}\|_* \ge \|\hat{\boldsymbol{X}}_{\mathrm{e}}\|_* = \|\boldsymbol{X}_{\mathrm{e}} + \boldsymbol{H}_{\mathrm{e}}\|_* \ge \|\boldsymbol{X}_{\mathrm{e}} + \mathcal{A}_{\Omega^\perp}(\boldsymbol{H}_{\mathrm{e}})\|_* - \|\mathcal{A}_{\Omega}(\boldsymbol{H}_{\mathrm{e}})\|_*. \tag{97}$$

The second term can be bounded using the triangle inequality as

$$\|\mathcal{A}_{\Omega}(\boldsymbol{H}_{\mathrm{e}})\|_{\mathrm{F}} \le \left\|\mathcal{A}_{\Omega}\left(\hat{\boldsymbol{X}}_{\mathrm{e}} - \boldsymbol{X}_{\mathrm{e}}^{\mathrm{o}}\right)\right\|_{\mathrm{F}} + \|\mathcal{A}_{\Omega}(\boldsymbol{X}_{\mathrm{e}} - \boldsymbol{X}_{\mathrm{e}}^{\mathrm{o}})\|_{\mathrm{F}}. \tag{98}$$

Since the constraint of Noisy-EMaC requires $\left\|\mathcal{P}_{\Omega}\left(\hat{\boldsymbol{X}} - \boldsymbol{X}^{\mathrm{o}}\right)\right\|_{\mathrm{F}} \le \delta$ and $\|\mathcal{P}_{\Omega}(\boldsymbol{X} - \boldsymbol{X}^{\mathrm{o}})\|_{\mathrm{F}} \le \delta$, the Hankel structure of the enhanced form allows us to bound $\left\|\mathcal{A}_{\Omega}\left(\hat{\boldsymbol{X}}_{\mathrm{e}} - \boldsymbol{X}_{\mathrm{e}}^{\mathrm{o}}\right)\right\|_{\mathrm{F}} \le \sqrt{n_1 n_2}\delta$ and $\|\mathcal{A}_{\Omega}(\boldsymbol{X}_{\mathrm{e}} - \boldsymbol{X}_{\mathrm{e}}^{\mathrm{o}})\|_{\mathrm{F}} \le \sqrt{n_1 n_2}\delta$, which immediately leads to

$$\|\mathcal{A}_{\Omega}(\boldsymbol{H}_{\mathrm{e}})\|_{\mathrm{F}} \le 2\sqrt{n_1 n_2}\delta.$$

Using the same analysis as for (62) allows us to bound the perturbation $\mathcal{A}_{\Omega^\perp}(\boldsymbol{H}_{\mathrm{e}})$ as follows

$$\|\boldsymbol{X}_{\mathrm{e}} + \mathcal{A}_{\Omega^\perp}(\boldsymbol{H}_{\mathrm{e}})\|_* \ge \|\boldsymbol{X}_{\mathrm{e}}\|_* + \frac{1}{4}\|\mathcal{P}_{T^\perp}\mathcal{A}_{\Omega^\perp}(\boldsymbol{H}_{\mathrm{e}})\|_{\mathrm{F}}.$$

Combining this with (97), we have

$$\|\mathcal{P}_{T^\perp}\mathcal{A}_{\Omega^\perp}(\boldsymbol{H}_{\mathrm{e}})\|_{\mathrm{F}} \le 4\|\mathcal{A}_{\Omega}(\boldsymbol{H}_{\mathrm{e}})\|_* \le 4\sqrt{n_1 n_2}\|\mathcal{A}_{\Omega}(\boldsymbol{H}_{\mathrm{e}})\|_{\mathrm{F}} \le 8 n_1 n_2 \delta.$$

Further from Lemma 3, we know that

$$\left\|\mathcal{P}_T\mathcal{A}_{\Omega^\perp}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\|_{\mathrm{F}} \leq \frac{n_1 n_2}{m}\sqrt{2}\left\|\mathcal{P}_{T^\perp}\mathcal{A}_{\Omega^\perp}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\|_{\mathrm{F}}. \tag{99}$$

Therefore, combining all the above results give

$$\begin{aligned}\|\boldsymbol{H}_{\mathrm{e}}\|_{\mathrm{F}} &\leq \|\mathcal{A}_\Omega(\boldsymbol{H}_{\mathrm{e}})\|_{\mathrm{F}} + \left\|\mathcal{P}_T\mathcal{A}_{\Omega^\perp}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\|_{\mathrm{F}} + \left\|\mathcal{P}_{T^\perp}\mathcal{A}_{\Omega^\perp}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\|_{\mathrm{F}} \\ &\leq \left\{2\sqrt{n_1 n_2} + 8 n_1 n_2 + \frac{8\sqrt{2} n_1^2 n_2^2}{m}\right\}\delta.\end{aligned}$$

K Proof of Theorem 4

In order to extend the results to structured Hankel matrix completion, from the proof of Theorem 1 it is sufficient to have the first two conditions in (38) to hold for general Hankel matrices. The proof is done by recognizing these two conditions are equivalent to (26).

References

[1] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.

[2] L. Potter, E. Ertin, J. Parker, and M. Cetin, "Sparsity and compressed sensing in radar imaging," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1006–1020, 2010.

[3] L. Borcea, G. Papanicolaou, C. Tsogka, and J. Berryman, "Imaging and time reversal in random media," *Inverse Problems*, vol. 18, no. 5, p. 1247, 2002.

[4] L. Schermelleh, R. Heintzmann, and H. Leonhardt, "A guide to super-resolution fluorescence microscopy," *The Journal of cell biology*, vol. 190, no. 2, pp. 165–175, 2010.

[5] A. M. Sayeed and B. Aazhang, "Joint multipath-Doppler diversity in mobile wireless communications," *IEEE Transactions on Communications*, vol. 47, no. 1, pp. 123 –132, Jan 1999.

[6] Y. Chi, Y. Xie, and R. Calderbank, "Compressive demodulation of mutually interfering signals," *submitted to IEEE Transactions on Information Theory*, 2013. [Online]. Available: http://arxiv.org/abs/1303.3904

[7] E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[8] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289 –1306, April 2006.

[9] E. J. CandÃšs, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.

[10] X. Li, "Compressed sensing and matrix completion with constant proportion of corruptions," *Constructive Approximation*, vol. 37, pp. 73–99, 2013.

[11] Y. Chi, L. Scharf, A. Pezeshki, and A. Calderbank, "Sensitivity to basis mismatch in compressed sensing," *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 2182–2195, May 2011.

[12] M. Duarte and R. Baraniuk, "Spectral compressive sensing," *Applied and Computational Harmonic Analysis*, 2012.

[13] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, "Compressed sensing off the grid," July 2012. [Online]. Available: http://arxiv.org/abs/1207.6053

[14] R. Prony, "Essai experimental et analytique," *J. de l'Ecole Polytechnique (Paris)*, vol. 1, no. 2, pp. 24–76, 1795.

[15] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 37, no. 7, pp. 984 –995, Jul 1989.

[16] Y. Hua and T. K. Sarkar, "Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 38, no. 5, pp. 814 –824, may 1990.

[17] D. Tufts and R. Kumaresan, "Estimation of frequencies of multiple sinusoids: Making linear prediction perform like maximum likelihood," *Proceedings of the IEEE*, vol. 70, no. 9, pp. 975 – 989, sept. 1982.

[18] Y. Hua, "Estimating two-dimensional frequencies by matrix enhancement and matrix pencil," *IEEE Transactions on Signal Processing*, vol. 40, no. 9, pp. 2267 –2280, Sep 1992.

[19] P. L. Dragotti, M. Vetterli, and T. Blu, "Sampling moments and reconstructing signals of finite rate of innovation: Shannon meets strang-fix," *IEEE Transactions on Signal Processing*, vol. 55, no. 5, pp. 1741 –1757, May 2007.

[20] M. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with finite rate of innovation," *IEEE Transactions on Signal Processing*, vol. 50, no. 6, pp. 1417–1428, 2002.

[21] K. Gedalyahu, R. Tur, and Y. C. Eldar, "Multichannel sampling of pulse streams at the rate of innovation," *IEEE Transactions on Signal Processing*, vol. 59, no. 4, pp. 1491–1504, 2011.

[22] E. Candes and Y. Plan, "A probabilistic and RIPless theory of compressed sensing," *IEEE Transactions on Information Theory*, vol. 57, no. 11, pp. 7235–7254, 2011.

[23] Y. Chi, A. Pezeshki, L. L. Scharf, and R. Calderbank, "Sensitivity to basis mismatch in compressed sensing," *ICASSP*, Mar. 2010.

[24] E. J. Candes and C. Fernandez-Granda, "Towards a mathematical theory of super-resolution," *to appear in Communications on Pure and Applied Mathematics*, 2013.

[25] ——, "Super-resolution from noisy data," November 2012. [Online]. Available: http://arxiv.org/abs/1211.0290

[26] V. Chandrasekaran, B. Recht, P. Parrilo, and A. Willsky, "The convex algebraic geometry of linear inverse problems," *48th Annual Allerton Conference on Communication, Control, and Computing*, pp. 699–703, 2010.

[27] E. J. Candes and B. Recht, "Exact matrix completion via convex optimization," *Foundations of Computational Mathematics*, vol. 9, no. 6, pp. 717–772, April 2009.

[28] D. Gross, "Recovering low-rank matrices from few coefficients in any basis," *IEEE Transactions on Information Theory*, vol. 57, no. 3, pp. 1548–1566, March 2011.

[29] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *Journal of ACM*, vol. 58, no. 3, pp. 11:1–11:37, Jun 2011.

[30] S. Negahban and M. Wainwright, "Restricted strong convexity and weighted matrix completion: Optimal bounds with noise," *The Journal of Machine Learning Research*, vol. 98888, pp. 1665–1697, May 2012.

[31] M. Wu, "Collaborative filtering via ensembles of matrix factorizations," vol. 2007, 2007.

[32] R. Salakhutdinov and N. Srebro, "Collaborative filtering in a non-uniform world: Learning with the weighted trace norm," *Advances in Neural Information Processing Systems (NIPS)*, 2010.

[33] T. Zhang, M. Alley, M. Lustig, X. Li, J. Pauly, and S. Vasanawala, "Fast 3D DCE-MRI with sparsity and low-rank enhanced SPIRiT (SLR-SPIRiT)," *Proceedings of the 21th Annual Meeting of ISMRM*, p. 2624, 2013.

[34] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Review*, vol. 52, no. 3, pp. 471–501, 2010.

[35] M. Fazel, T. K. Pong, D. Sun, and P. Tseng, "Hankel matrix rank minimization with applications in system identification and realization," 2011.

[36] I. Markovsky, "Structured low-rank approximation and its applications," *Automatica*, vol. 44, no. 4, pp. 891–909, 2008.

[37] B. Balle and M. Mohri, "Spectral learning of general weighted automata via constrained matrix completion," *Advances in Neural Information Processing Systems (NIPS)*, pp. 2168–2176, 2012.

[38] A. Sankaranarayanan, P. Turaga, R. Baraniuk, and R. Chellappa, "Compressive acquisition of dynamic scenes," *Computer Vision–ECCV 2010*, pp. 129–142, 2010.

[39] P. J. Shin, P. E. Larson, M. A. Ohliger, M. Elad, J. M. Pauly, D. B. Vigneron, and M. Lustig, "Calibrationless parallel imaging reconstruction based on structured low-rank matrix completion," *submitted to Magnetic Resonance in Medicine*, 2012.

[40] M. Fazel, H. Hindi, and S. P. Boyd, "Log-det heuristic for matrix rank minimization with applications to Hankel and Euclidean distance matrices," *American Control Conference*, vol. 3, pp. 2156 – 2162 vol.3, June 2003.

[41] J. F. Cai, E. J. Candes, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.

[42] M. Grant, S. Boyd, and Y. Ye, "Cvx: Matlab software for disciplined convex programming," *Online accessiable: http://stanford. edu/˜ boyd/cvx*, 2008.

[43] N. Alon and J. H. Spencer, *The Probabilistic Method (3rd Edition)*. Wiley, 2008.

[44] E. Candes and Y. Plan, "Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements," *IEEE Transactions on Information Theory*, vol. 57, no. 4, pp. 2342–2359, 2011.